DOCUMENT RESUME

ED 179 578 TM 009 859

AUTHOR Cohen, Allan S., Comp.
TITLE Bibliography of Papers on Latent Trait Assessment.
INSTITUTION Educational Testing Service, Evanston, Ill.
SPONS AGENCY Office of Education (DHEW), Washington, D.C.
PUB DATE 2 Feb 79
NOTE 64p.

EDRS PRICE MF01/PC03 Plus Postage.
DESCRIPTORS Attitude Tests; Computer Programs; Equated Scores; *Evaluation Methods; Factor Analysis; Goodness of Fit; Item Analysis; *Mathematical Models; *Measurement; Program Evaluation; Sampling; *Scores; Scoring; *Statistical Analysis; Technical Reports; Testing; Test Reliability; Tests
IDENTIFIERS Computer Assisted Testing; Elementary Secondary Education Act Title I; *Latent Trait Theory; Rasch Model; Test Theory

ABSTRACT

This partially annotated bibliography of journal articles, dissertations, convention papers, research reports, and a few books and unpublished manuscripts provides a comprehensive coverage of work on latent trait theory and practice. Documents are arranged alphabetically by author. The period covered ranges from the early 1950's to the present. The bibliography was prepared to assist state and local education agencies make appropriate use of latent trait models in the evaluation of their Title I programs. (Author/CTM)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED179578

U S DEPARTMENT OF HEALTH,
EDUCATION & WELFARE
NATIONAL INSTITUTE OF
EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

BIBLIOGRAPHY OF PAPERS ON LATENT TRAIT ASSESSMENT

compiled by

Allan S. Cohen

Region V Technical Assistance Center
Educational Testing Service
Midwestern Regional Office
One American Plaza
Evanston, Illinois 60201

IM009 859

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

ALLAN S. COHEN

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

EDUCATIONAL TESTING SERVICE
PRINCETON, NEW JERSEY

PURPOSE OF THIS BIBLIOGRAPHY

This is a specialized bibliography of references pertaining to latent trait models. It was developed as part of the activities of Region V Technical Assistance Center (TAC) associated with the U. S. Office of Education's evaluation of Title I programs.

The bibliography was prepared to assist state and local education agencies make appropriate use of latent trait models in the evaluation of their Title I programs. Annotations of references were included (and will be added) so as to maximize the usefulness of this compilation for those agencies.

So as to remain current, the bibliography will be continually updated throughout the year. Copies of new references and/or additional annotations to existing references will be available unbound the first of each quarter (i.e., March 1, July 1, and October 1). (Requests for additional bound copies may be made at any time during the year and will be filled until the supply is exhausted). Region V TAC will likewise endeavor to make available as many of the non-public references contained within this bibliography as is possible. Papers which are journal publications, laboratory reports, research bulletins, etc., will not be provided.

Although this bibliograpphy was prepared in conjunction with a U.S.O.E. contract, Region V TAC is solely responsible for the accuracy of its contents.

BIBLIOGRAPHY OF PAPERS ON

LATENT TRAIT ASSESSMENT

compiled by

Allan S. Cohen

Region V Technical Assistance Center
Educational Testing Service
Midwestern Regional Office
One American Plaza
Evanston, Illinois 60201

February 2, 1979

BIBLIOGRAPHY OF PAPERS ON
LATENT TRAIT ASSESSMENT**

Educational Testing Service
Midwestern Regional Office
One American Plaza
Evanston IL 60201

Ace, M. E. Application of a population-free scaling technique to attitude data. Unpublished doctoral dissertation, University of Minnesota, 1972. (Forster, Passmore)

Ace, M. E., & Barth, R. T. The Rasch Model in attitude measurement. Proceedings of the 81st Annual Convention of the American Psychological Association, Montreal, 1973, 8, 47-48. Explored the fit between real attitude data and the Rasch model. The study was based on limited research showing that the model can provide objective ability measurement, and recent suggestive results of applying it to attitude data. 3 sets of 7 point unipolar rating scale items, with at least 350 respondents each, were evaluated, using indices of fit for the overall data and for single items. Even with deletion of deviant items in subsequent iterations, results were largely negative. It is concluded that attitude data may not be appropriate for use with the Rasch model. (Forster, Passmore, Roberts)

Andersen, E. B. The discrete measurement model. Copenhagen: University of Copenhagen Press, 1964 (available in Danish only). (Forster, Passmore)

Andersen, E. B. Posterior analysis of individual parameters in item analysis. (Report No. 4) Copenhagen, Denmark: Copenhagen School of Economics and Business Administration, 1968. (Forster, Passmore)

Andersen, E. B. Asymptotic properties of conditional maximum likelihood estimators. Journal Royal Statistical Society, Series B, 1970, 32, 283-301. (Forster, Passmore)

Andersen, E. B. A strictly conditional approach in estimation theory. Skandinavish Akturietidskrift, 1971, 137-152. (Forster, Passmore)

Andersen, E. B. The asymptotic distribution of conditional likelihood ratio tests. Journal of the American Statistical Association, 1971, 66 (335), 630-633. (Forster, Passmore)

*Indicates articles that are available from ETS-MRO.

**Contributors to this bibliography are noted in the parentheses following references thay have provided.

Andersen, E. B. A computer program for solving a set of conditional maximum likelihood equations arising in the Rasch model for questionnaires. Research Memorandum 72-6. Princeton, N. J.: Educational Testing Service, 1972. A computer program for solving the conditional likelihood equations arising in the Rasch model for questionnaires is described. The estimation method and the computational problems involved are described in a previous research report by Andersen, but a summary of those results are given in two sections of this paper. A working example is also provided to help the reader understand different computational steps of the program; these include (1) defining properties of the model, (2) computing estimates of the item parameters by taking the conditional distribution of the total set of responses given the individual score vectors as the likelihood function, (3) applying a goodness-of-fit test based on a grouping of possible score vectors, (4) developing technical, input-output and main subroutines. Numerical examples are given of two actual executions of the program. The appendix, comprising nearly half of the report, lists the main program and subroutines. (Forster, Passmore)

Andersen, E. B. The numerical solution of a set of conditional estimation equations. The Journal of the Royal Statistical Society, Series B, 1972, 34, 42-54. (Forster, Passmore)

Andersen, E. B. A goodness of fit test for the Rasch Model. Psychometrika, 1973, 38, 123-140. Discuses the Rasch model, an item analysis model with logistic item characteristic curves of equal slope, i.e., with constant item discriminating powers. The proposed goodness-of-fit test is based on a comparison between difficulties estimated from different scoregroups and over-all estimates. Based on the within score group estimates and the over-all estimates of item difficulties a conditional likelihood ratio is formed. It is shown that -2 times the logarithm of this ratio is x2-distributed when the Rasch model is true. The power of the proposed goodness-of-fit test is discussed for alternative models with logistic item characteristic curves but unequal discriminating powers of the items. The developed theory is illustrated by an analysis of 20 items from a scholastic aptitude test. (Forster, Passmore, Roberts)

Andersen, E. B. Conditional inference and models for measuring (Publikation Nr. 5). Copenhagen, Denmark: Mentalhygiejnisk Forskningsinstitut, 1973. (Forster, Passmore)

Andersen, E. B. Conditional inference for multiple-choice questionnaires. British Journal of Mathematical and Statistical Psychology, 1973, 26, 31-44. (Forster, Passmore)

Andersen, E. B. Sufficient statistics and latent trait models. Psychometrika, 1977, 42, 69-81. For questionnaires with two answer categories. It has been proven in complete generality that if a minimal sufficient statistic exists for the individual parameter and if it is the same statistic for all values of the item parameters, then the raw score (or the number of correct answers) is the minimal

sufficient statistic. It follows that the model must be of the Rasch type with logistic item characteristic curves and equal item-discriminating powers. This paper extends these results to multiple choice questionnaires. It is shown that the minimal sufficient statistic for the individual parameter is a function of the so-called score vector. It is also shown that the so-called equidistant scoring is the only scoring of a questionnaire that allows for a real valued sufficient statistic that is independent of the item parameters, if a certain ordering property for the sufficient statistic holds. (Roberts)

Andersen, E. B, & Madsen, M. Estimating the parameters of the latent population distribution. Psychometrika, 1977, 42(3), 357-74. Methods for estimating the mean and variance of latent ability parameters of a normally distributed population that has been tested with Rasch model-calibrated test items are discussed. Methods for checking the normality of the population are also included.

Anderson, J., Kearney, G. E. & Everett, A. V. An evaluation of Rasch's structural model for test items. British Journal of Mathematical and Statistical Psychology, 1968, 21, 231-238. (Forster, Passmore, Roberts)

Anderson, L. W. Measuring change in academic self-concept resulting from curricular and instructional innovations. Journal of Youth and Adolescence, 1977, 6(3), 215-228.

Andrich, D. Latent trait psychometric theory in the measurement and evaluation of writing ability. Unpublished doctoral dissertation, University of Chicago, 1973.

Andrich, D. Binomial latent trait model for study of Likert-style attitude questionnaires. British Journal of Mathematical and Statistical Psychology, 1978, 31, 84-98.

Angoff, W. H. Scales, norms, and equivalent scores. In R. L. Thorndike (Ed.), Educational Measurement (2nd ed.). Washington D.C.: American Council on Education, 1971. (Forster, Passmore)

*Arneklev, B., Geis, D., & Ingebo, G. Optimum range of difficulty for linking items. In C. Karr (Chair), Exploring the Rasch test scaling procedure. Symposium presented at meeting of the American Educational Research Association, San Francisco, 1976. One of the most important contentions of the Rasch model of item analysis is that two tests of the same trait, having some items in common, can be linked together using a "linking constant" derived from the common items. This would be accomplished by administering both tests to a sample of testees, calibrating the items of the tests independently, and then linking the test by a linking constant which would be the difference between the average calibration of the common items on test one and the independently determined average calibration of the common items on test two. An important question regarding this linking procedure is whether the accuracy of the link is affected

significantly when the items used for the linking represent a restricted range of difficulty relative to all items in the tests being linked. This paper empirically addresses this question with a variety of anaylses using test results from mathematics tests administered to samples of fourth through seventh grade pupils. While the analyses varied with respect to the specific questions that they addressed, the overall conclusion was that for these tests, restriction in range of item difficulty did not present serious difficulties for test linking using the Rasch procedures.

Baker, F. B. Automation of test scoring, reporting, and analysis. In R. L. Thorndike (Ed.), Educational Measurement (2nd ed.). Washington, DC: American Council on Education, 1971. (Forster, Passmore)

Baker, F. B. Advances in item analysis. Review of Educational Research, 1977, 47, 151-178. Recent advances in item analysis have provided greater capabilities for the analysis of tests, but have also significantly increased the gap between the theory and practice of item analysis. This paper traces the lines of development in item analysis under latent trait theory.

Bart, W. M., & Airasian, P. W. An empirical examination of the relationship between test factor structure and test hierarchical structure. Paper presented at meeting of the American Educational Research Association, San Francisco, 1976. The question of whether test factor structure is indicative of the test item hierarchy was examined. Data from 1,000 subjects on two sets of five bivalued Law School Admission Test items, which were analyzed with latent trait methods of Bock and Lieberman and of Christoffersson in Psychometrika, were analyzed with an ordering-theoretical method to locate item hierarchies. Though one item set was unifactor and the other bifactor, both item sets showed no hierarchies even after 10 percent of the subjects responsible for low response pattern frequencies were deleted. Factor structure of a test appears to reveal nothing about the test's hierarchical structure.

Bartenwerfer, H. Measurement of social political attitudes using Rasch scale. Psychologische Rundschau, 1977, 28(3), 209.

Bashaw, W. L. Measurement objectivity and homogeneity. Paper presented at the Annual Meeting of the National Council on Measurement in Education, Minneapolis, 1970. (Forster, Passmore)

Bashaw, W. L, Rentz, R. R. & Wright, B. D. An invitational conference on the Rasch model and its applications to standardized test data. Sponsored by Harcourt, Brace, & Jovanovich, New York City, February 4, 1971. (Forster, Passmore)

Bech, P., Allerup, P., & Rosenberg R. Marke-Nymann temperament scale - evaluation of transferability using Rasch item analysis. Acta Psychiatrica Scandinavica, 1978, 57(1), 49-58.

Bejar, I. I. An investigation of the dichotomous, graded and continuous response level latent trait models. (Doctoral dissertation, University of Minnesota, 1975). Dissertation Abstracts International, 1975, 36(7-B), 3566.

Benson, J. G. A comparison of three types of item analysis in test development using classical and latent trait methods. (Doctoral dissertation, University of Florida, 1977). Dissertation Abstracts International, 1977, 39(1-A), 243.

Berkson, J. A statistically precise and relatively simple method of estimating the bioassay with quantal response, based on the logistic function. Journal of the American Statistical Association, 1953, 48, 565-599.

Berkson, J. Maximum likelihood and minimum x^2 estimates of the logistic function. Journal of the American Statistical Association, 1955, 50, 130-162.

Betz, N. E., & Weiss, D. J. An empirical study of computer-administered two-stage ability testing. Research Report 73-4. Minneapolis, MN: University of Minnesota, Department of Psychology, Psychometric Methods Program, October, 1973.

Beyerlein, G., & Hike, R. Factor analysis and G. Rasch's measuring model: Procedures in scale construction and the possibilities of interpretation in evaluating educational goals, as indicated by two empirical investigations. Friedrich-Alexander-U. Zeitschrift fur Experimentelle und Angewandte Psychologie, 1975, 22(3), 453-470. Compared the factor-analyzed results of questionnaire items with an evaluation of the same data using G. Rasch's individual-centered approach to item analysis. It remains questionable whether dimensions derived by factor analysis can be used for scale construction and subsequent analysis with the G. Rasch model. The importance of explicit content-theoretical considerations in the selection of items was clearly established. (Roberts)

Birnbaum, A. Classification by ability levels. In F. M. Lord & M. R. Novick, Statistical theories of mental test scores. Reading, MA: Addison-Wesley, 1968. (Forster, Passmore)

Birnbaum, A. Some latent trait models and their use in inferring an examinee's ability. In F. M. Lord & M. R. Novick, Statistical theories of mental test scores. Reading, MA: Addison-Wesley, 1968. (Forster, Passmore)

Birnbaum, A. Statistical theory for logistic mental test models with a prior distribution of ability. Journal of Mathematical Psychology, 1969, 6, 258-276. Research Bulletin 67-12, Princeton, NJ: Educational Testing Service, 1967.

Black, R. H. A method of discriminating partial knowledge. (Doctoral dissertation, University of Oklahoma). Dissertation Abstracts International, 1975, 36(1-A), 181.

Blommers, P. A ratio scale for test item difficulty. Unpublished manuscript, University of Iowa, 1965. (Forster, Passmore)

Bock, R. D. Estimating item parameters and latent ability when responses are scored in two or more nominal categories. Psychometrika, 1972, 37, 29-51. Proposes a multivariate logistic latent trait model for items scored in 2 or more nominal categories. Statistical methods based on the model provide (a) estimation of 2 item parameters for each response alternative of each multiple choice item, and (b) recovery of information from "wrong" responses when estimating latent ability. An application to a large sample of data from 557 Ss for 20 vocabulary items showed excellent fit of the model according to a chi-square criterion. Item and test information curves are compared for estimation of ability assuming multiple category and dichotomous scoring of these items. Multiple scoring proves substantially more precise for Ss of less than median ability, and about equally precise for Ss above the median.

Bock, R. D. Basic issues in the measurement of change. In D. N. de-Gruitjer and L. J. Van der Kamp (Eds.), Advances in psychological and educational measurement. London: Wiley and Sons, 1976.

Bock, R. D., & Thrash, W. Characterizing a latent trait distribution. Proceedings of the 1976 Dayton Symposium on Applications of Statistics, Dayton, 1976.

Bock, R. D., & Wood, R. Test theory. In P. H. Mussen and M. E. Rosenzweig (Eds.), Annual Review of Psychology, 1971, 22, 193-224. Reviews the literature on test theory from 1966-1969. Discussion is limited to "test theory as it has developed around problems of measuring ability and attainment by means of objective tests." Areas include general and classical test theory; strong true score theory; latent trait theory; other related topics (i.e., reliability theory, item sampling and generalizability theory, decision-theoretic considerations, item construction theory, item scoring theory, and time-related models); and computer procedures. Speculation on the future of these methods in applied settings is made. (Forster, Passmore)

Boldt, R. F. An estimation procedure for the Rasch model allowing missing data. Research Memorandum 72-5, Princeton, N. J.: Educational Testing Service, 1972. The Rasch model and other latent trait models encounter some difficulty when faced with an appreciable amount of missing data or omitting behavior. The present note assumes that some reasonable missing data model has been formulated which does not involve the parameters associated with the latent ability of interest. A maximum likelihood function is used that is based on probabilities which are conditional on the occurrence of a response. (Forster, Passmore)

Boruch, R. F. A procedure for estimating the amount of trait, method, and error variance attributable to a measure. Unpublished doctoral dissertation, Iowa State University, 1968.

Boruch, R. F., & Wolins, L. A procedure for estimation of trait, method, and error variance attributable to a measure. Educational and Psychological Measurement, 1970, 30, 547-574.

Bostrup, L., & Prien, B. The basis of the mathematical-statistical analysis of the subtests in the Danish revision of the Binet test. Skolepsykologi, 1970, 7(4), 263-278. Describes the model for item analysis introduced by G. Rasch. It is a probabilistic model with 2 sets of parameters, e.g., task and person parameters, based on test items with only right-wrong solutions (a more extended version has been developed). The person parameters for 2 or more subtests may be functions of the same person parameter and the question of a total result or profile is to be examined empirically. The task parameter need not correspond to the Ss feelings about difficulty. (Roberts)

Bradshaw, C. W. The use of invariant item parameters to derive an absolute score metric. (ERIC Document Reproduction Service No. ED 150 171). A method for determining invariant item parameters is presented, along with a scheme for obtaining test scores which are interpretable in terms of a common metric. The method assumes a unidimensional latent trait and uses a three parameter normal ogive model. The assumptions of the model are explored, and the methods for calculating the proposed statistics are appended.

Bramble, W. J. A least squares method of parameter estimation of the logistic measurement model. Unpublished manuscript, University of Kentucky, 1970. (Forster, Passmore)

Brigman, S. L. Rasch model test equating with three multiple test equating designs. (Doctoral dissertation, University of Georgia). Dissertation Abstracts International, 1976, 37(6-A), 3570.

*Brigman, S. L., & Bashaw, W. L. Multiple test equating using the Rasch Model. Paper presented at the meeting of the American Educational Research Association, San Francisco, 1976. Procedures are presented equating simultaneously several tests which have been calibrated by the Rasch model. Three multiple test equating designs are described. A Full Matrix Design equates each test to all others. A Chain Design links tests sequentially. A Vector Design equates one test to each of the other tests. For each design, the Rasch model test equating constants were obtained for four reading vocabulary tests. The standard errors of the constants based on each design are also provided and the appropriate use of each design is discussed.

Brink, N. E. Characteristics of Rasch's logistic model. Paper presented at the meeting of the American Educational Research Association, Minneapolis, MN, 1970. (Eric Document Reproduction Service No. ED 042 804). A basic description of the Rasch model and a brief review of the work previously done on the model is related. Simulated data was used to test goodness of fit to the Rasch model. It was found that data with near perfect fit to the Guttman model provides perfect fit to the Rasch model, though a perfect Guttman scale collapses.

Random data also provides a good fit. The effects of varying the standard deviations on normally distributed total scores and the ranges on uniformly distributed scores and the effects of combining two sets of independent data are analyzed in some detail. Not available in hard copy doe to marginal legibility of original document. (Forster, Passmore)

Brink, N. E. The effect of item discrimination and range of item easiness on the standard error of ability estimates using the Rasch model. (Doctoral dissertation, Univerity of California, Los Angeles) Dissertation Abstracts International, 1971, 31(8-A), 3947. (Forster, Passmore)

Brink, N. E. Overview of the Rasch model. Proceedings of the 79th Annual Convention of the American Psychological Association, 1971b, 6, (1), 95-96. Presents topics that tie the Rasch model to other more commonly understood topics, thus clarifying its position in the area of measurement and research. This model produces ability estimates that are independent of the sample of Ss and the set of items used, thus it is a latent trait model. The simplicity of this 2-parameter model allows this objectivity or independence within the model to occur. Yet this makes the model more restrictive as to the characteristics of items that compose the test, items need to be of equal discrimination. Because of the advantages of objectivity, refinement of items to fit this model seems called for. (Forster, Passmore, Roberts)

Brink, N. E. Effect of item discrimination in the Rasch model. Proceedings of the Annual Convention of the American Psychological Association, 1971, 6(Pt. 1) 101-102. Item discrimination is a parameter not provided for in the Rasch model. The effect of this parameter in measuring ability was exarined using simulated data. Two effects were noted: (a) increases in item discrimination decreases the unit of measure of the ability, (b) increases in item discrimination decreases the number of items that contribute to the measurement of a particular level of ability thus increasing the standard error of that level of activity. Increases in the density of items at points along the ability continuum decrease the standard error of ability estimates at those points. (Forster, Passmore, Roberts)

Brink, N. E. Rasch's logistic model vs. the Guttman model. Educational and Psychological Measurement, 1972, 32(4), 921-927. Study compares the Rasch and the Guttman models of measurement and thus adds to the description of the characteristics of Rasch's logistic model. Such knowledge is of importance in making decisions as to which model and which statistics should be used in evaluation of tests.

Brogden, H. E. The Rasch model, the law of comparative judgment and additive conjoint measurement. Psychometrika, 1977, 42, 631-634. Relationships between the Rasch test analysis model, the law of comparative judgment, and additive conjoint measurement are discussed. (Roberts)

Brooks, R. D. An empirical investigation of the Rasch ratio-scale model for item-difficulty indexes (Doctoral dissertation, State University of Iowa). Dissertation Abstracts International, 1965, 26(04), 2047. (University Microfilms No. 65-00434) (Forster, Passmore)

Brown, S. H. A comparison of the accuracy of conventional tests constructed and scored using the Rasch, Urry, and classical test models. (Doctoral dissertation, University of Minnesota). Dissertation Abstracts International, 1977, 38(6-B), 2920.

Burns, J. A. Comparison of classical and Rasch model approaches to the measurement of group differences. (Doctoral dissertation, Northwestern University). Dissertation Abstracts International, 1977, 37(8-A), 5062.

Butollo, W. H. Logistic test model approach to changes of factorial structure during learning. Journal of General Psychology, 1972, 86(2), 189.

Cartledge, C. M. A comparison of equipercentile and Rasch equating methodologies. (Doctoral dissertation, University of Georgia). Dissertation Abstracts International, 1975, 37(8-A), 5141.

Choppin, B. An item bank using sample-free calibration. Nature, 1968 219 (5156), 870-872. (Forster, Passmore)

Choppin, B. Recent developments in item banking. In D. N. de Gruitjer & L. J. VanderKamp (Eds.), Advances in Psychological and Educational Measurement. London: Wiley, 1976.

Clark, C. L. (Ed.), Proceedings of the First Conference on Computerized Adaptive Testing. Washington, D.C.: United States Civil Service Commission, 1976. The principal objectives of this conference were to exchange information, discuss theoretical and empirical developments, and to coordinate research efforts. The papers and their authors are: "The graded response model of latent trait theory and tailored testing" by F. Samejima; "Incomplete orders and computerized testing" by N. Cliff; "Adaptive testing research at Minnesota: Over view, recent results, and future directions" by D. J. Weiss; "Adaptive testing research at Minnesota: Some properties of a Bayesian sequential adaptive mental testing strategy" by J. R. McBride; "An empirical investigation of Weiss' stradaptive testing model" by B. K. Waters; "Using computerized tests to add new diensions to the measurement of abilities which are important for on-job performance: An exploratory study" by C. H. Cory; "A broad-range tailored test of verbal ability" and "Some likelihood functions found in tailored testing" by F. M. Lord; "Bayesign tailored testing and the influence of item bank characteristics" by C. J. Jensema; "Reflections on adaptive testing" by D. N. Hansen; "Computer-assisted testing: An orderly transition from theory to practice" by R. H. testing feasible?" by V. W. Urry; "Effectiveness of the ancillary estimation procedure" by J. F. Gugel, F. L. Schmide, & V. N. Urry; and "Item parameterization procedures for the future" by F. L. Schmidt and V. W. Urry. (ERIC Document Reproduction Service No. ED 126 110)

Cohen, L. A. A modified logistic response model for item analysis. Unpublished manuscript, 1976.

Colonius, H. On Keats' generalization of the Rasch model. Psychometrika, 1977, 42, 443-445. Parameter estimation for Keats generalization of the Rasch model that takes account of guessing behavior is investigated. It is shown that no minimal sufficient statistics for the ability parameters independent of the difficulty parameters exist. (Roberts)

Computers and testing: Steps toward the inevitible conquest. Symposium held at the meeting of the American Psychological Association, 1975. This pamphlet reprints three papers and an invited discussion of them, read at a Division 5 Symposium at the 1975 American Psychological Association Convention. The first paper describes a Bayesian tailored testing process and shows how it demonstrates the importance of using test items with high discrimination, low guessing probability, and a uniform (rectangular) distribution of difficulty. The second paper describes a relatively easy, inexpensive, and reliable method of estimating these three item parameters. The procedure assumes that the underlying ability is normally distributed. Data were examined to determine the reasonableness of the assumption. The third paper reports the results of two simulation studies evaluating the effectiveness of the parameterization technique, showing that the procedure works well under "standard" conditions (at least 2000 cases and 100 items) and is adequate with 1000 cases and 60 items. The discussion paper points out that accurate estimations of item parameters are essential to successful tailored testing; compares the Bayesian technique with parameterization based on maximum likelihood estimates; and concludes that there is now no apparent obstacle to successful operational tailored testing. (ERIC Document Reproduction Service No. ED 135 853).

Connolly, A. J., Nachtman, W., & Pritchett, E. M. Key Math Diagnostic Arithmetic Test. Circle Pines, Minn.: American Guidance Service, 1971, 1974. (Forster, Passmore)

Conrad, W., et. al. Construction of Rasch scales based on the anxiety questionnaires FS 5-10 and KAT. Diagnostica, 1976, 22(3), 110-125. Two anxiety scales were subjected to a scaling procedure based on G. Rasch's model (1960), which is stronger than the classical test theory model and yields scales with more desirable properties. Results confirm that the items of scales constructed on the basis of classical criteria tend to fit the stronger Rasch model. A number of reasons are given to support the view that a desirable scale construction strategy is to select items initially on the basis of classical criteria (item difficulty, point-biserial item-scale correlation), and then subject the items which survive this preliminary testing to a more refined and complex selection procedure based on Rasch. (Roberts)

Conrad, W., et. al. Contributions to the problem of the metric of subjective personality as represented by the E and N scales of the HANES, KJI. Diagnostica, 1976, 22(10, 13-26. Estimated the item and person

parameters of the logistic model of G. Rasch (1960) from the responses of 2,224 pupils in Grades 5-10 to the items of a measure of neuroticism and extraversion for children and adolescents. For each of the 2 personality dimensions, a scale which fit the Rasch model was constructed. The traditional item statistics of these scales were satisfactory, indicating that traditional methods of personality test construction based on point-biserial item-total scale correlations are likely to yield scales which fit the Rasch model to some degree. (Roberts)

Cormier, G. An investigation of the fit of the Rasch measurement model to data from the medical college admission test. (Doctoral dissertation, Michigan State University). Dissertation Abstracts International, 1977, 38(5-A), 2723.

Cudeck, R. A., Cliff, N., Reynolds, T. J., & McCormick, D. J. Monte Carlo results from a computer program for tailored testing. (Technical Report No. 2) Los Angeles: University of Southern California, Department of Psychology, February, 1976.

Curry, A. R. Invariance of the Rasch model ability estimates over different collections of items. (Doctoral dissertation, University of Georgia). Dissertation Abstracts International, 1977, 38(11-A), 6664.

Cypress, B. K. The effects of diverse test score distribution characteristics on the estimation of the ability parameter of the Rasch measurement model. (Doctoral dissertation, Florida State University). Dissertation Abstracts International, 1972, 33(6-A), 2761. (Also paper presented at the meeting of the American Educational Research Association, New Orleans, 1973). The potential of the Rasch model to develop scores, on a ratio scale, suitable for inter-individual comparisons, from intact groups with disparate distribution charactgeristics was investigated. The specific problems studied were: (1) the effects of skewed test score distributions on the ability parameter of the Rasch measurement mode; (2) the effects of group size on the ability parameter of the Rasch measurement model; (3) the interactive effects of skewed test score distributions and group size on the ability parameter of the Rasch measurement model; (4) the effects of skew and total group size on the standard errors of estimate of item log easiness estimates. The data for the study in the form of item responses were randomly selected from 120,000 students who participated in the Florida State-Wide Testing Program in September 1971. The 90-item mathematics test was selected. Thirty-five raw score distributions characterized by seven levels of skew and five group sizes were constructed. Group responses were submitted to a computer program which estimated the model's parameters according to a maximum likelihood procedure. Results of the study indicated that the estimates derived from the Rasch measurement model were not independent of the group used to produce them. Differences were minimal in the middle score range, but large in low and high score range. Eleven tables present the study data. (Forster, Passmore)

Dahl, M. Program MAIN: Maximum likelihood estimation of the ability parameter of Rasch's model from a set of previously calibrated items. Unpublished computer program, American Guidance Service, Circle Pines, MN, 1972. (Forster, Passmore)

Denham, C. H. Criterion-referenced, domain-referenced and norm-referenced measurement - parallax view. Educational Technology, 1975, 15(12), 9-13.

Dinero, T. E., & Haertel, E. A computer simulation investigating the applicability of the Rasch model with varying item discriminations. Paper presented at the meeting of the National Council on Measurement in Education, San Francisco, 1976. This paper will discuss the results of a series of computer simulations comparing the Rasch logistic model to a series of models departing to various degrees from its assumption of equal discrimination power for all items. The results have implications for test construction and test scoring, indicating how closely the conventional raw score corresponds to the mathematically correct score, in which each item response is weighted by an index of the item discrimination.

Docherty, E. M. A test of Piaget's qualitative developmental hypothesis using the Rasch model. (Doctoral dissertation, Syracuse University). Dissertation Abstracts International, 1973, 34(3-8), 1255-1256.

Doenges, D. I., & Scheller, R. Scaling Rorschach responses with the one-dimensional Rasch model. U. Trier, W. Germany. Psychologische Beitrage, 1976, 18(3), 390-409. Argues that less than satisfactory research using the Rorschach as a diagnostic tool is related to the failure to provide an adequate psychometric analysis of responses to the test stimuli. It is also argued that, of the 3 recognized approaches to the quanitification of psychological variables, response scaling is most adequate for coping with the intricate nature of the Rorschach testing situation. This situation is likely to be more amenable to description with probabilistic than with deterministic models. In order to assess the validity of this conclusion, responses of 933 psychiatric patients were analyzed with the model developed by G. Rasch (1960). Regularities postulated by this model were found to hold for 3 Rorschach variables, and results point to the practicability of a latent trait approach to scaling the Rorschach. (Roberts)

Doherty, V. W. Developments in goal based measurement in the Portland public schools. Paper presented at the meeting of the American Educational Research Association, San Francisco, 1976. For almost 20 years, the Portland public schools have maintained a testing program anchored in locally constructed tests. The district's newly developed tests are called goal-referenced tests, a version of criterion-referenced tests, that meet the district's purposes because they reflect what well-qualified teachers in the district believe should be taught and measured. In constructing the tests,

curriculum personnel select goals they believe the test should measure; teachers develop items that measure those goals; the tests are given a trial administration and items are analyzed; the items, formats, and directions are revised; and the tests are administered for their intended use. The Rasch test scaling procedure that involves the identification of an equal-interval scale of difficulty for a given set of items based on information about item difficulty and total test performance for the group tested is used. The procedure yields information on item difficulty and an estimate of the ability of individuals and groups tested. This permits establishment of a scale that is independent of the norming population and allows for the creation of item pools.

Doherty, V. W. Domain referenced measurement using a Rasch calibrated item bank (Occasional Paper No. 18). Portland, Oregon: Portland Public Schools, 1978.

Doherty, V. W., & Forster, F. Can Rasch scaled scores be predicted from a calibrated item pool? In G. Karr (Chair), Exploring the Rasch test scaling procedure. Symposium presented at the meeting of the American Educational Research Association, San Francisco, 1976. In order to verify that a test made of items selected from a pool produces the same Rasch scaled achievement scores as would be predicted from previous Rasch difficulty calibrations for those items, the following experiment was conducted. Seven math tests with similar content but graduated difficulty were administered to students in grade seven. The tests overlapped each other so that each test included about 20 items that were also included in one of the other six tests. The items in these seven tests were calibrated for difficulty using the Rasch procedure. Through the linking data provided by the common items, each of the items in all of the tests was adjusted to the same scale of difficulty. These items then formed a single pool of items. These items were listed in ascending difficulty order and were divided into seven level tests each with 30 items. These new tests were administered to about 1500 students in the seventh grade. Rasch item calibrations were again computed, and linking data used to adjust items to a common achievement scale. Rasch scaled achievement scores were computed for each possible raw score for each of the level tests and for the same items as calibrated in the original pool. The comparable scaled achievement scores for the tests were established on the basis of the actual performance of students on the second administration.

Douglas, G. A. Test design strategies for the Rasch psychometric model. (Doctoral dissertation at the University of Chicago). Dissertation Abstracts International, 1976, 36(7-A) 4427.

Douglas, G. A. Conditional maximum-likelihood estimation for a multiplicative binominal response model. British Journal of Mathematical and Statistical Psychology, 1978, 31, 73-83.

Draba, R. E. The Rasch model and legal criteria of a "reasonable" classification. (Doctoral dissertation, University of Chicago). Dissertation Abstracts International, 1978, 39(1-A), 245.

Durovic J. Application of the Rasch Model to civil service testing. Paper presented at the meeting of the Northeastern Educational Research Association, Grossingers, New York, 1970. The New York State Department of Civil Service investigated an empirical application of the Rasch model which appears useful in Civil Service testing. The model is a powerful tool for developing insights into what test items are measuring while permitting the investigator to spot defective items. It also reveals meaningful distinctions in the type of task set by different items. Two sets of specific examples are discussed to illustrate its usefulness. The first set of examples considers the use of item probability as an index of the degree of fit of the material to the model, while the second discusses the "normal deviate" matrix, which displays the goodness of fit of each item at each score group, and enables an investigator to ascertain the overall validity of the general index. These examples demonstrate the applicability of the Rasch model to a variety of conditions. The author suggests that the model seems promising for civil service testing since it is not simply a means to derive scores but also a powerful tool for test analysis, construction, and design. (Forster, Passmore)

Durovic J. Improving written test measurement in personnel selection. Public Personnel Review, in press. (Forster, Passmore)

Durovic, J. Test bias: An objective definition for test items. Paper presented at the meeting of the Northeastern Educational Research Association, 1975. A test bias definition, applicable at the item-level of a test is presented. The definition conceptually equates test bias with measuring different things in different groups, and operationally equates test bias with a difference in item fit to the Rasch model, greater than one, between groups. It is suggested that the proposed definition avoids confusing etiology with measurement by capitalizing on the "objectivity" property of the logistic Rasch measurement model. Application of the definition, to 914 applicants (black = 367; white = 547) in a "real" selection situation is described and resulted in identifying two items as biased. The two items so defined, were different than the two items identified as biased by comparing the item success rates (i.e., item difficulty) of black vs. white candidates. A content evaluation of the items by two black, female reviewers was subsequently performed. Their comments lend preliminary support to the proposed psychometric test bias definition. Additional encouraging support is provided by the match between the content comments and the item bias index values, for other items in the test. Implications for future applications and research are presented.

Elliott, C. D., & Murray D. J. Measurement and speed of problem-solving and its relation to childrens age and ability. British Journal of Educational Psychology, 1977, 47, 50-59.

Elton, C. F., & Rose, H. A. A vocational interest test minus sex bias. Journal of Vocational Behavior, 1975, 7(2), 207-214. The Vocational Preference Inventory responses from 290 subjects were subjected to a Rasch item analysis, one of a class of latent trait models. After elimination of 22 items which did not fit the model, a sex-free form of the VPI was obtained.

English, R. A., Reckase, M. D., & Patience, W. M. Application of tailored testing to achievement measurement. Behavior Research Methods & Instrumentation, 1977, 9(2), 158-161. Compared differences on the levels of achivement, differences in time needed to take a test, and differences in attitudes between Ss taking a traditional paper-pencil test and those taking a computerized tailored test. A maximum likelihood tailored testing procedure based on the Rasch simple logistic model was used with 108 upper level undergraduate students enrolled in a measurement and evaluation course. Analyses include descriptive statistics on all variables for all conditions, intercorrelations of all variables within each treatment condition, and analyses of variance. Results indicate no differences in the level of achievement between Ss taking paper-pencil tests and those taking the tailored tests. A time savings of 75% was found using the tailored testing model. Ss indicated that tailored tests were less stressful. (Roberts)

Epstein, K. I. An empirical investigation of four criterion-referenced testing models. Paper presented at the conference of the Military Testing Association, 1975. Since the primary purpose of classical testing is to rank order examinees consistently, the absolute value of the true score has been relatively unimportant. However, the major purpose of criterion referenced testing is to estimate the true capabilities of examinees to perform specific tasks. Hence, the problems of true score determination assume critical importance. Four measurement models which have potential for evaluating the results of criterion referenced tests are discussed here. The proportion correct model assumes that the proportion of sample trials scored correct is an unbiased estimate of the proportion correct in the infinite domain for that individual. The binomial error model adds the specification of the conditional distribution for the observed score for the given true proportion correct. The third model applies the philosophy implied in the binomial error model to Bayesian statistical theory, and the final model is the Rasch one parameter logistic model. Advantages and disadvantages of each are discussed, but the final choice of a model is to be based on the needs of the testing program and the resources available to analyze the data.

Farr, S. D. What is the Rasch Model? Paper presented at the Annual Meeting of the Northeastern Educational Research Association, Crossingers, New York, November, 1970. (Forster, Passmore)

Ferris, M. J. Validity as a function of empirical scaling of test items by a logistic model. Educational and Psychological Measurement, 1967, 27, 829-835. In general, the findings tend to indicate that rescoring tests by using item weights computed by the logistic model does not improve the predictive validity of a test. Still, since the correlation between weights with the raw criterion scores was slightly higher than the correlation between raw predictor scores and the raw criterion scores, there may be merit in using the logistic model to determine indices of difficulty of test items for improving predictive validity. (Roberts)

Fhaner, S., & Sjoberg, L. Measurement in differential psychology: A symposium. Goteborg Psychological Reports, 1973, 3(6), 42 pages. Summarizes papers on (a) latent trait theory, including conditional estimation in the Rasch model and dimensionality in a latent trait model; (b) generalizability of behavioral data, including the construct validity of tests and empirical studies in differential psychology; and (c) measurement of change, including models and estimation methods and path models with latent variables. (Roberts)

Fischer, G. H. Psychologische Testtheorie. Bern, Germany: Huber, 1966. (Forster, Passmore)

Fischer, G. H. A further note on estimation in Rasch's measurement model with two categories of answers. Research Bulletin No. 3. Vienna, Austria: The University of Vienna, 1970. (Forster, Passmore)

Fischer, G. H. Data models and parameter models. Zeitschrift fur experimentelle und angewandte Psychologie, 1970, 17(2), 212-219. Discusses 3 approaches to measurement based on: (a) preestablished rules, (b) deterministic data models, and (c) probabilistic parameter models. The last approach is the most promising since it leads to concept formation and theory building. As an example, the application of Rasch's logistic measurement model on Rorschach and Holtzman tests is given.

Fischer, G. H. A measurement model for the effect of mass media. Acta Psychologica, 1972, 36, 207-220. (Forster, Passmore)

Fischer, G. H. Conditional maximum likelihood estimation of item parameters for a linear logistic test model. Research Bulletin No. 9, Vienna, Austria: The University of Vienna, 1972. (Forster, Passmore)

Fischer, G. H. A step towards dynamic test theory. Research Bulletin No. 10. Vienna, Austria: The University of Vienna, 1972. (Forster, Passmore)

Fischer, G. H. Some probalistic models for measuring change. In D. DeGruijter & L. I. van der Kamp (Eds.), Advances in Psychological and Educational Measurement. London: Wiley, 1976.

Fischer, G. H., & Formann, A. K. An algorithm and a FORTRAN program for estimating the item parameters of the linear logistic test model. Research Bulletin No. 11. Vienna, Austria: The University of Vienna, 1972. (Forster, Passmore)

Fischer, G. H., & Kower, P. New traffic perception test on basis of logistic test model. Accident Analysis and Prevention, 1977, 9(3), 203-215.

Fischer, G. H., & Scheiblechner, H. Algorithms and programs for the probabilistic test model of Rasch. Psychologische Beitrage, 1970, 12(1), 23-51. Developed 2 procedures for parameter estimation of G. Rasch's 1-dimensional test model. Both procedures were applied to empirical and simulated data. Satisfactory convergence and precision of estimation were achieved. A test was developed to find out whether the model is useful in the case of a special set of data. All procedures were programmed in FORTRAN IV. Program descriptions and lists of programs and sub-programs are included. (Roberts)

Fischer, G. H., & Spada, H. F. The metric analysis of a projective form interpretation test. Psychologische Beitrage, 1970, 12(1), 65-82. Investigated the question of whether and in what way quantitative statements on personality traits can be made by means of results of interpretation tests, on the basis of 305 Holtzman test protocols. The 1-dimensional test of G. Rasch helped to isolate those test variables which could be used for setting up dichotomies of answer categories which corresponded to the assumptions of the model. It is suggested that those scoring variables which do not meet the demands of the model be eliminated. (Roberts)

*Forbes, D. W. The use of Rasch logistic scaling procedures in the development of short multi-level arithmetic achievement tests for public school measurement. In C. Karr (Chair), Exploring the Rasch test scaling procedure. Symposium presented at the meeting of the American Educational Research Association, San Francisco, 1976. (ERIC Document Reproduction Service No. ED 128 400) Rasch calibration permitted the development of short achievement tests that were economical in testing time, and could be developed in a series of difficulty levels to suit student individual differences. Furthermore these tests were of adequate reliability for practical educational measurement when indivual students were assigned to tests of appropriate difficulty level. A variety of test placement strategies were considered and several were tried. Two formal procedures involving the use of a pretest screening tool for level assignment show promise of effectiveness but in the research described tended to place many children in a test which was somewhat too difficult for them. The use of screening tests still is considered very promising although it is recommended that in the future criteria for test placement be

modified so the students would be placed one, or perhaps two, levels lower than they were in the field test of these prototypes. It is further recommended that any students who get raw scores under 5, or over 25, immediately be retested with a more appropriate level to forestall the dramatic measurement error increases which occur when those limits are exceeded.

Forbes, D. W., & Ingebo, G. S. An empirical test of the content homogeneity assumption involved in Rasch item calibration. Paper presented at the meeting of the American Educational Research Association, Washington, D. C., 1975. A project was carried out to determine the degree of content homogeneity that a test item pool must have in order to accomplish successful Rasch calibration. Mathematics item pools were administered to upper elementary children. The items were analyzed under two conditions, with items organized into separate subtests and as a global mathematics test. Results indicate that satisfactory item calibration can be carried out with global mathematics tests, avoiding the necessity of organizing items into separate subtests. The degree to which calibration is identical under the two conditions is a topic for further study.

*Forster, F. The optimum size for Rasch field testing (How big is enough?) Unpublished manuscript. Portland, Oregon: Portland Public Schools, 1975 (Forster)

*Forster, F. Sample size and stable calibrations. In C. Karr (Chair), Exploring the Rasch test scaling procedure. Symposium presented at the meeting of the American Educational Research Association, San Francisco, 1976. (Forster)

*Forster, F. Straight talk about the Rasch. Unpublished manuscript. Portland, Oregon: Portland Public Schools, 1976. (Forster)

Forster, F. The Rasch item characteristic curve and actual item performance. In C. Karr (Chair), Exploring the Rasch test scaling procedure. Symposium presented at the meeting of the American Educational Research Association, San Francisco, 1976. (ERIC Document Reproduction Service No. ED 129 840) Various factors which influence the relationship between the Rasch item characteristic curve and the actual performance of an item are identified. The Rasch item characteristic curve is a new concept in test design and analysis. The Rasch test model provides information concerning the percent of students with a specified achievement level who would be expected to correctly answer a question with a specified difficulty. Using this information, it is possible to make a continuous plot of the expected percent correct across the full range of achievement. This plot constitutes the Rasch item characteristic curve. From the data gathered in this study it appears that using a minimum score group size of three to five students, a refined fit cutoff of .05 to .10 are optimum. Further research will be directed toward cross validating these findings. (Forster)

*Forster, F. What the columns on the Rasch printout mean. Unpublished manuscript. Portland, Oregon: Portland Public Schools, 1976. (Forster)

*Forster, F. Historical documentation: Development of the multi-level 7th grade math tests. Unpublished manuscript. Portland, Oregon: Portland Public Schools, 1976. (Forster)

*Forster, F. The theory of the Rasch model. Unpublished manuscript. Portland, Oregon: Portland Public Schools, 1977. (Forster)

*Forster, F. The application of the Rasch (For school testing folk). Unpublished manuscript. Portland, Oregon: Portland Public Schools, 1977. (Forster)

*Forster. F. Everything you wanted to know about the Rasch. Unpublished manuscript. Portland, Oregon: Portland Public Schools, 1978. (Forster)

*Forster, F., & Ascher, G. The Rasch calibrated item bank: A new tool for competency based evaluation. Unpublished manuscript. Portland, Oregon: Portland Public Schools, Oregon State Department of Education, no date. (Forster)

*Forster, F., & Ascher, G. Research of the Rasch measurement model. Unpublished manuscript. Portland Oregon: Portland Public Schools, Oregon State Department of Education, 1977. (Forster)

*Forster, F., & Doherty, V. Using the Rasch measurement model to the advantage in designing a school testing program. Unpublished manuscript. Portland, Oregon: Portland Public Schools, 1977. (Forster)

*Forster, F., & Doherty, V. Monograph II: Can the Rasch Difficulty Level of an item be determined without relying on random sampling? Unpublished manuscript. Portland, Oregon: Portland Public Schools, 1977. (Forster)

*Forster, F., & Doherty, V. Monograph III: Is the Rasch item level scale equal interval? Unpublished manuscript. Portland, Oregon: Portland Public Schools, 1977. (Forster)

*Forster, F., & Doherty, V. Monograph IV: Do students receive the same Rasch achievement levels on different tests? Unpublished manuscript. Portland, Oregon: Portland Public Schools, 1977. (Forster)

*Forster, F., & Doherty, V. Monograph V: What is the best sample size for field testing? Unpublished manuscript. Portland, Oregon: Portland Public Schools, 1977. (Forster)

*Forster, F., & Doherty, V. Monograph VI: Is the calibration of an item level affected by the range of items levels on the test? Unpublished manuscript. Portland, Oregon: Portland Public Schools, 1977. (Forster)

*Forster, F., & Doherty, V. Monograph VII: The mathematics of the Rasch model. Unpublished manuscript. Portland, Oregon: Portland Public Schools, 1977. (Forster)

*Forster, F., & Doherty, V. Using the Rasch approach to measurement to solve practical school testing problems. Paper presented at the meeting of the American Educational Research Association, Toronto, March 1978. (Forster)

*Forster, F., & Ingebo, G. S. Linking. Unpublished manuscript. Portland, Oregon: Portland Public Schools, no date. (Forster)

*Forster, F., & Ingebo, G. S. Research on the Rasch measurement model. (Occasional Paper No. 13). Portland, Oregon: Portland Public Schools, 1978. (Forster)

*Forster, F., & Ingebo, G. S. Measurement sense and crition nonsense. (Occasional Paper No. 16). Portland, Oregon: Portland Public Schools, 1978. (Forster)

*Forster, F., & Ingebo, G. S. Reporting student achievement using a learning based scale. Paper presented at meeting of the Oregon Educatinal Research Association, 1978. (Forster)

Fowler, E. P., & Bramble, W. J. An analysis of personality data using a Rasch measurement model. A paper presented at the meeting of the American Educational Research Association, Chicago, April 1972. The applicability of the Rasch model to data from a typical personality test, the High School Personality Questionnaire (HSPQ), was studied. The data were gathered on Junior High and High School students in the Louisville Public Schools (Kentucky). Item easinesses and person abilities were estimated and compared by age group, within each age group, and across two points in time for the older age group. In addition, certain results from the Rasch analyses were compared with those of factor analysis. A sample of 1,000 students was taken from each of the groups (Junior High, 7th graders; Senior High, 11th and 12th graders). Results of the study are related to five questions considered. The first question was whether or not there were patterns of fit to the Rasch model when responses are dichotomized in different ways. The results indicated that no single key was superior to others in producing fit. The second question was concerned with fit of the model for the data considered; it was found that frequently there was lack of fit, but it is noted that the test statistic was conservative. The third question related to the stability of item easinesses estimates within a group

and across two points in time for that group. The conclusion was that different item easinesses are obtained when different degrees of possession of the trait are focused upon. The fourth question was how stable the tests of fits results are across time; pre- and post-comparisons of fit found 55% in agreement. The fifth question concerned how the item mean squares are related to factor loadings; in almost all cases, the item with the highest mean square was also the item with the lowest loading. (Forster, Passmore)

Fryman, J. G. Application of the Rasch simple logistic model to a mathematics placement examination. (Doctoral dissertation, University of Kentucky). Dissertation Abstracts International, 1976, 37(09-A), 5626.

Garrison, W. M. The use of the Rasch measurement model in developmental self-concept research. (Doctoral dissertation, Purdue University). Dissertation Abstracts International, 1976, 38(2-A), 702.

*Garrison, W. M., Cohen A. S., & Linden, K. W. The use of the Rasch measurement model in developmental analyses of children's self-concept. Paper presented at the meeting of the American Educational Research Association, New York, April 1977. The utility of the Rasch logistic measurement model in longitudinal analyses of children's self-concept was investigated. Calibration of an 80- item measure from the responses of 1,000 elementary school children resulted in the identification of 25 items that fit the model. Repeated measures analyses of variance used to assess the effects of an experimental intervention program on changes in individuals' self-concept tended to support the usefulness of the measure in reducing the number of test items needed to measure the latent variable in question. Some evidence on the claim for item subtest equivalence also as obtained.

Goldstein, H., & Blinkhorn, S. Monitoring educational standards - inappropriate model. Bulletin of the British Psychological Society, 1977, 30, 309-311.

Gomez, H. The Cali nutrition, health and educational project: technical report. Cali, Columbia: Human Ecology Research Station, 1973.

Gomez, H. The analysis of growth. Unpublished doctoral dissertation, Northwestern University, 1978.

*Gomez, H. Evaluating longitudinal data with the use of the Rasch model. Unpublished research report, Northwestern University.

Green, B. J. Comments on tailored testing. In W. H. Holtzman (Ed.), Computer-assisted instruction, testing, and guidance. New York: Harper & Row, 1970.

Gugel, J. F., Schmidt, F. L., & Urry, V. W. Effectiveness of the ancillary estimation procedure. Proceedings of the First Conference on Computerized Adaptive Testing (PS-75-6, Personnel Research and Development Center, U. S. Civil Service Commission). Washington, D. C.: U. S. Government Printing Office, 1976. (Superintendent of Documents Stock No. 006-000-00940-9).

Gullikse, H. Looking back and ahead in psychometrics. American Psychologist, 1974, 29(4), 251-261.

Gundelach, P., & Tetzschner, H. Measurement of influence in organizations - critique of control-graph method. Acta Sociologica, 1976, 19(1), 49-63.

*Gusky, T. R. Use of the Rasch model in calibrating the Reading Comprehension Test of the Iowa Tests of Basic Skills. Chicago, Illinois: Chicago Board of Education, Department of Research and Evaluation, 1978.

Gustafsson, J. E. The Rasch model for dichotomous items: Theory, Applications and a computer program. No. 63. Gothenburg, Sweden: Gothenberg University, Institute of Education, 1977. The Rasch model for test analysis is described and compared with two-parameter and three-parameter latent-trait models. Conditional maximum likelihood equations for estimating item parameters are derived, and estimates of person parameters are described together with their confidence intervals. Goodness of fit tests are discussed, including a graphic test of item fit and two over-all tests. Characteristics of tests which may cause them not to fit the model are listed, together with strategies for developing tests. Applications of the Rasch model to optimizing test efficiency, test equating and linking, and to tailored testing are described. Some generalizations of the basic model and special cases are mentioned. PML, A FORTRAN IV computer program used in appolying the Rasch model, is described in detail, and a number of bibliographical references are appended.

Haberman, S. J. Maximum likelihood estimates in exponential response models. Annals of Statistics, 1977, 5(5), 815-841.

Haladyna, T. M. The quality of domain-referenced test items. Paper presented at the meeting of the American Educational Research Association, San Francisco, April, 1976. The objectives of this study were to first determine whether or not the empirical item analysis of domain referenced tests (DR) was justified; and second, in the event that it was, which of a set of recommended procedures was most effective for determining item quality. The analysis that followed led to the conclusion that empirical procedures were highly desirable. When these empirical procedures were applied to test data, the results indicated that four different techniques provided almost identical information: Rasch statistics, instructional sensitivity indexes, traditional statistics, and Baysian indexes. Based on these results, it would seem that any one of these four would serve adequately.

Haladyna, T. M., & Roid, G. A theoretical and empirical comparison of three approaches to achievement testing. (ERIC Document Reproduction Service No. ED 148 903). Three approaches to the construction of achievement tests are compared: construct, operational, and empirical. The construct approach is based upon classical test theory and measures an abstract representation of the instructional objectives. The operational approach specifies instructional intent through instructional objectives, facet design, item forms, and item-generating algorithms. Criterion referenced and domain referenced tests are discussed, as they relate to the operational approach. A series of trials and subsequent calibrations are used in the empirical approach, which is based upon latent trait theory. To compare these three approaches, an achievement test was administered to 258 dental students, before and after instruction in basic dental anatomy. Five different strategies were used to construct the test forms: random selection, item-test correlations, pre-post test difficulty index, Bayesian, and Z-difference. Descriptive statistics were computed, including estimated and actual reliability, and estimated and true standard errors of measurement. The results suggested that a random sampling approach provides the greatest representativeness and minimizes errors of measurement. Random sampling is recommended for both the construct and operational definition approaches to measurement. A completely valid test of the third approach has not yet been completed.

Hambleton, R. K. Computation of information curves and efficiency for logistic test models: a Fortran IV program for the CDC 3600 computer. Amherst, MA: University of Massachusetts, Center for Educational Research, 1970. (Forster, Passmore)

Hambleton, R. K. An empirical investigation of the Rasch test theory model (Doctoral dissertation, University of Toronto, 1969). Dissertation Abstracts International, 1971, 31 (08-A), 4035 (Available from Author, University of Massachusetts, School of Education, Amherst, Mass. 01002) (Forster, Passmore)

Hambleton, R. K. Contributions to criterion-referenced test theory: On the uses of item characteristic curves and related concepts. Research Report No. 51. Amherst, MA: University of Massachusetts, School of Education, Laboratory of Psychometric and Evaluative Research, 1977.

Hambleton, R. K., & Cook, L. L. Latent trait models and their use in the analysis of educational test data. Journal of Educational Measurement, 1977, 14(2), 75-96. Latent trait models offer considerable potential for the improvement of educational measurement practice, but until recently they have received only limited attention from measurement specialists. Reasons for the limited attention include the complexity of the models, their relative newness, the shortage of technical guidelines for their use, and some doubts about the size of the gains that might accrue from their use. This paper has

three primary purposes: First, to provide a non-mathematical introduction to latent trait models and some of their features. Specifically it discusses assumptions, shapes of item characteristic curves, ability scales, parameter estimation, and the availability of computer programs to conduct latent trait analyses. Second, it provides a comparison of classical test models and latent trait models. Finally, it introduces two promising applications of latent trait models.

Hambleton, R. K., & Novick, M. R. Toward an integration of theory and method for criterion-referenced tests. Journal of Educational Measurement, 1973, 10, 159-170.

Hambleton, R. K., & Rovinelli, R. A. FORTRAN IV program for generating examinee reponse data from logistic test models. Behavioral Science, 1973, 18, 74.

Hambleton, R. K., Swaminathan, H., Cook, L. L., Eignor, D. R., & Gifford, J. A. Developments in latent trait theory: A review of models, technical issues, and applications. A paper presented at the joint meeting of the National Council on Measurement in Education and the American Educational Research Association, New York, April 1977. (Eric Document Reproduction Service No. ED 137 364). Latent trait theory supposes that, in testing situations, examinee performance on a test can be predicted (or explained) by defining examinee characteristics, referred to as traits, estimating scores for examinees on these traits and using the scores to predict or explain test performance (Lord and Novick, 1968). In view of the breakthroughs in several testing problem areas brought about by the use of latent trait theory, it is clear that the field of latent trait theory will become increasingly more important to measurement specialists and test practitioners. This paper comprehensively reviews this field and addresses four matters. First, the nature and characteristics of latent trait theory are introduced. Second, a review of many of the technical developments in the field is provided. Third, several promising applicatins of latent trait models are described. Finally, some additional areas for research and development are suggested.

Hambleton, R. K., & Traub, R. E. Information curves and efficiency of three logistic test models. British Journal of Mathematical and Statistical Psychology, 1971, 24, 273-281. (Also, Technical Report No. 4. Amherst, MA: University of Massachusetts, Center for Educational Research, 1970 and Proceedings of the 78th Annual Meeting of the American Psychological Association, 1970, 5, 121-122). (Forster, Passmore, Roberts)

Hambleton, R. K., & Traub, R. E. Analysis of empirical data using two logistic latent trait models. British Journal of Mathematical and Statistical Psychology, 1973, 26, 195-211. Georg Rasch has developed a new one-parameter latent trait model to explain the performance of examinees on achievement tests. The model can be viewed as a special case of Birnbaum's two-parameter logistic model where all items are assumed to have equal discriminating power. Both models assume that

guessing does not occur. This study compares how well the one- and two-parameter models fitted different sets of empirical data. For each model, a measurement of agreement was made between the expected and obtained distributions of ability estimates using three sets of data. The results revealed that the more general the model, the better the fit with real data. (Forster, Passmore)

Hambleton, R. K., & Traub, R. E. The robustness of the Rasch test model. (Report No. 42). Amherst, MA: University of Massachusetts, Laboratory of Psychometric and Evaluative Research, 1976. (Forster, Passmore)

Hamm, D. W. Estimates of item characteristics for the dichotomous and multiplicative binomial Rasch models. Paper presented at the meeting of the American Educational Research Association, New York, 1977.

Harris, M. L., & Romberg, T. A. An analysis of content and task dimensions of methematics items designed to measure level of concept attainment. Journal for Research in Mathematics Education, 1974, 5(20, 72-86. Used factor analytic techniques to study content and task dimensions of mathematics items, developed to measure concept attainment, using a crossed design involving 30 concepts and 12 tasks. Conventional factor analyses were performed separately for 196 female 5th graders and 195 male 6th graders for concept and task scores. Also, 3-mode factor analyses were performed. From results of the conventional factor analyses, it is concluded that all 30 of the concepts and measures of a single functional relationship existing among the concepts and that all 12 tasks are measures of a single underlying ability or latent trait. The 3-mode results indicate that there are no important concept-task interactions for idealized persons. Thus, it is reasonable to regard the concepts and tasks as being 2 independent modes.

Hashway, R. M. A comparison of tests derived using Rasch and traditional psychometric paradigms. (Doctoral dissertation, Boston College, 1977.) Dissertation Abstracts International, 38(2A), 744.

Hashway, R. M. Individual and program evaluation using the Rasch model. New York: Praeger Publishers, 1978. A clear and concise manual for practitioners, administrators and program specialists concerned with understranding and interpreting Rasch test scores, an item selection technique for use in standardized tests. Because of the difficulty of interpreting Rasch model test scores, the use of this method is not as widespread as the use of test item selections based on percentile ratings or on the number of correct responses. To correct this problem, the author presents a critical review of the Rasch model to date and provides a model of Rasch test scores in terms of true scores and measurement error. He then offers procedures designed to compare Rasch scores and to convert them into familiar scales such as T-scores. Numerous examples are supplied to amplify theory and to illustrate the methods involved. (Publisher's Abstract)

Hehl, F. J., & Nussell, E. Qualitative and quantitative differentiation of complaints using Rasch model. Zeitschrift fur Klinische Psychologie und Psychotherapie, 1975, 23(4), 331-342.

Heimann, H. Methodological problems of clinical trials of psychotropic drugs. Archiv fur Psychiatrie und Nervenkrankheiten, 1975, 220(3), 255-268. Argues that clinical studies have only succeeded in differentiating between great categories of psychotropic drugs, but failed to prove finer differences of effects within these categories of substances. 2 points of the testing method are discussed: (a) problems which arise when rating pathological behavior and (b) problems of sampling psychiatric patients. It is noted that many of the symptoms that clinicians and psychologists formerly considered as relevant were extremely rare, and it is maintained that even in clinical studies, it would be possible to choose among the variety of descriptive symptoms those which fulfill requirements of the probabilistic test-model of Rasch and to take only those symptoms to characterize the degree of severity of psychic disturbance in trials with psychotropic drugs. The main problems of sampling are discussed, i.e., the loss of information as a consequence of taking the mean in a group of psychiatric patients, and the role of biological rhythms. It is demonstrated, using 2 cases of depressive patients, that an enormous difference of psychophysiological responsiveness can be hidden behind very similar clinical pictures. (Roberts)

Heinrich, I. Changes in syllogistic reasoning influenced by semantic variation of propositions. Psychologische Beitrage, 1975, 17(4), 497-518.

Henning, H. J. Mokken scale analysis. Psychologische Beitrage, 1976, 18(3), 410-430.

Henning, H. J., & Six, B. Structure of Machiavellianism-scale. Zeitschrift fur Sozialpsychologie, 1977, 8(3), 185-198.

Herrmann, T., & Stapf, K. Marburg 2-component model of parental reinforcement - remarks on critique by Lukesch. Zeitschrift fur Entwicklungspsychologie und Padagogische Psychologie, 1977, 9(3), 198-204.

Herrmann, T., Stapf, A., & Krohne, H. W. The Marburg scales of parent-child relations. Diagnostica, 1971, 17(3), 118-131. Reports construction of 2 sets of parent-child relations scales based on a 2-dimensional model of parental influence. The conceptual dimensions are parental severity and parental support. Evidence pertaining to the validity of the model is reviewed in detail. The 1st set of scales was constructed factor-analytically; the 2nd was based on the items of the 1st, consists of 4 15-item scales, and is based on the stochastic measurement model of G. Rasch (see PA, Vol. 36:5HB84R). The 4 final scales (severity of mother, severity of father, support of mother, support of father) are listed. Administered to a sample of 176 pupils they obtained reliabilities ranging from .80 to 88. (Roberts)

Nommers, W. The psychometrics of decisions of children in risky behavior. Psychologische Beitrage, 1977, 19(10, 84-99. In previous investigations several personality variables of children were measured on the basis of behavior in decisive situations involving 3 risky alternatives. The logistic model is examined by means of 2 data records regarding (a) the selection of those variables allowing model-adequate assessment, and (b) the scaling of the behavior of the Ss. The polychotomic model seemed to meet all requirements by allowing the measuring of 2 personality variables, probability orientation and expectation orientation, by multi-categorical scaling of decision behavior. The connection between parameter assessment and situative traits is examined. In both data records many relations were found, thus confirming the portfolio theory.

Horst, P. Determination of optimal test length to maximize the multiple correlation. Psychometrika, 1949, 14, 79-88.

Horst, P. Optimal test length for maximum differential prediction. Psychometrika, 1950, 21, 51-66.

Immisch, P. Experimentation for anxiety reduction in children during school teaching. School Psychology, 1972, 18(3), 300-308.

*Ingebo, G. S. Item pool linking procedures. In G. Karr (Chair), Exploring the Rasch test scaling procedure. Symposium presented at the meeting of the American Educational Research Association, San Francisco, 1976.

*Ingebo, G. S. How to link tests to form an item pool. Unpublished manuscript. Portland, Oregon: Portland Public Schools, no date. (Forster)

*Ingebo, G. S. The Rasch model and C.B.E. Public School Realities. (Occasional Paper No. 14). Portland, Oregon: Portland Public Schools, 1978.

Ireland, C. M. An application of the Rasch one parameter logistic model to individual intelligence testing in a tailored testing environment. (Doctoral dissertation, University of Missouri, 1977). Dissertation Abstracts International, 1977, 37(9-A), 5766.

Jaspars, J. M. Dogmatic thinking: An inquiry into the meaning of direct and indirect attitude measurement. Gewin, 1971, 12(5), 309-329. Criticizes the usual interpretation of discrepancy between behavior and (tested) attitude. It is argued that the absence of a correlation between behavior and attitude does not imply that words and deeds do not relate to the same latent trait. It is proposed that the prediction validity of attitude measures will not be improved by using indirect techniques, but rather by operationalizing the concept of "attitude." However, indirect measures can be important tools for studying the theoretical status of the concept "attitude." A preliminary study of indirect measurement of an attitude (pro-Vietnam, anti-American) is reported. Language: Dutch.

Jenkins, G. P., & Wright, B. D. Taxation of income of multinational corporations - case of United States petroleum industry. Review of Economics and Statistics, 1975, 57(1), 1-11.

Jensema, C. H. An application of latent trait mental test theory to the Washington pre-college testing battery. (Doctoral dissertation, University of Washington). Dissertation Abstracts International, 1972, 33(1-B), 423.

Jensema, C. J. An application of latent trait mental test theory. British Hournal of Mathematical and Statistical Psychology, 1974, 27(1), 29-48. Uses empirical and Monte Carlo data to find solutions for problems in applying A. Birnbaum's (1968) 3-parameter logistic mental test model through Bayesian tailored testing.

Jensema, C. J. The validity of Bayesian tailored testing. Educational and Psychological Measurement, 1974, 34, 757-766.

Jensema, C. J. An application of latent trait mental test theory. British Journal of Mathematical and Statistical Psychology, 1976, 29, 183-213.

Jensema, C. J. Bayesian tailored testing and the influence of item bank characteristics. Proceedings of the First Conference on Computerized Adaptive Testing (PS-75-6, Personnel Research and Development Center, U. S. Civil Service Commission). Washington, D. C., U. S. Government Printing Office, 1976. (Superintendent of Documents Stock No. 006-000-00940-9).

Jensema, C. J. A simple technique for estimating latent trait mental test parameters. Educational and Psychological Measurement, 1976, 36, 705-715. A simple and economical method for estimating initial parameter values for the normal ogive or logistic latent trait mental test model is outlined. The accuracy of the method in comparison with maximum likelihood estimation is investigated through the use of Monte-Carlo data.

*Karr, C. Research report on the Rasch Model of test development. In C. Karr (Chair), Exploring the Rasch test scaling procedure. Symposium presented at the meeting of the American Educational Research Association, San Francisco, 1976. (Forster)

Kearns, J., & Meredith, W. Methods for evaluating empirical Bayes point estimates of latent trait scores. Psychometrika, 1975, 40(1), 373-384. Examines the question of how large a sample must be in order to produce empirical Bayes estimates which are preferable to other commonly used estimates, such as proportion correct observed score.

Kearney, G. E. Some aspects of the general cognitive ability of various groups of educational significance as measured by the Queensland Test. Unpublished doctoral dissertation, University of Queensland, Australia, 1966. (Forster, Passmore)

Kearney, G. E. Some aspects of the general cognitive ability of various groups of aboriginal Australians as assessed by the Queensland Test. Unpublished doctoral dissertation, University of Queensland, Australia, 1966. (Forster, Passmore)

Kearney, G. E., Anderson, J., & Everett, A. V. An evaluation of the RO3 Test using Rasch's structural model for test items. Research Report No. 13, Brisbane, Australia: 21st Psychological Unit, 1967. (Forster, Passmore)

Keats, J. A. Test theory. Annual Review of Psychology, 1967, 18, 217-238. (Forster, Passmore, Roberts)

Keats, J. A. Applications of projective transformations to test theory. Psychometrika, 1974, 39, 359-360. The theory of projective transformation is applied to the Rasch model for representing test performances. (Roberts)

Keesling, J. W. Computer programming of the model. Paper presented at an American Educational Research Association pre-session on Person-Free Item Calibration and Item-Free Person Measurement, Los Angeles, 1969. (Forster, Passmore)

Keesling, J. W. Sample-free test calibration--evaluation of fit-control of the model. Paper presented at the American Educational Research Association pre-session on Person-Free Item Calibration and Item-Free Person Measurement, Los Angeles, 1969. (Forster, Passmore)

Keller, M. Development of role-taking ability - social antecedents and consequences for school success. Human Development, 1976, 19(2), 120-132.

Kempf, W. F. Probabilistic models in experimental psychological test situations. Psychologische Beitrage, 1972, 14(1), 16-37. Generalizes the multidimensional test model of G. Rasch from 2 to N parameters. The generalized model allows application to experimental psychological situations where it may be desirable to measure contributions from S characteristics, instruments being used, and specific testing conditions. (Roberts)

Kempf, W. F. Evaluation of factor analysis as a psychological method. Psychologische Beitrage, 1972, 14(4), 610-625.

Kifer, E., & Bramble, W. The calibration of a criterion-referenced test. Paper presented at the meeting of the American Educational Research Association, Chicago, 1974. (ERIC Document Reproduction Service No. ED 091 334) A latent trait model, the Rasch, was fitted to a criterion-referenced test. Approximately 80 percent of the items fit the model. Those items which fit the model were then calibrated. Based on the item calibration, individual ability estimates and the standard error of those estimates were calculated. Using the ability estimates, it was possible, given any criterion level, to compute the probability of a person exceeding the criterion. The calibration was done on the final examination of a performance-based introductory educational psychology course.

Kolakowski, D. A behaviour-genetic analysis of spatial ability utilizing latent-trait estimation. Unpublished doctoral dissertation, University of Chicago, 1971.

Kolakowski, D. Latent trait estimation: Theory vs. practice. Paper presented at American Psychological Association Symposium on the New Psychometrics, 1972. Empirical results are presented as regards the implementation of a latent-trait psychometric model by means of conditional maximum likelihood estimation. Items are scored polychotomously into varying numbers of nominal categories and the test and item characteristic curves and information functions are examined. It is concluded that scoring items in four or more categories, as opposed to the usual dichotomous scoring, can increase information gain by a factor of two or more in the lower range of ability. Thus, the error of measurement is decreased to an extent equivalent to doubling the test length in this range. Alternatively, one can sample the range of ability in the target population with far fewer items. This latter property addresses itself directly to the empirical constraits on time and resources which are encountered in psychological testing.

Kolakowski, D., & Bock, R. A. A FORTRAN IV program for maximum likelihood item analysis and test scoring: Normal ogive model. (Research Memo No. 12). Chicago: University of Chicago, Educational Statistics Laboratory, 1970.

Kolen, M. J., & Whitney, D. R. Accuracy of estimating two parameter logistic latent trait parameters and implications for classroom tests. Paper presented at the meeting of the American Educational Research Association, Toronto, March 1978. (Kolen)

Krause, B. Scaling models in psychodiagnosis. Zeitschrift fur Psychologie, 1977, 185(2), 257-287.

Kuncel, R. B. Response processes and relative location of subject and item. Educational and Psychological Measurement, 1973, 33(3), 545-563.

Kuncel, R. B., & Fiske, D. W. Stability of response process and response. Educational and Psychological Measurement, 1974, 34(4), 743-755.

Kuncel, R. B. Subject-item interaction in item-metric research. Educational and Psychological Measurement, 1977, 37(3), 665-678. The interaction of subjects with test items is investigated. It is suggested that psychometricians rely too heavily on inferences about the nature of the interaction. An approach to data analysis is proposed which is more directly related to this interaction.

Lei[illegible], D. J., & Weiss, D. J. An empirical comparison of two-stage and [illegible]ability testing. (Research Report 75-1). [illegible]ersity of Minnesota, Department of Psychology, Psychometric Methods Program, February 1975.

Lenke, J. M. The application of the Rasch Model to achievement test data. Paper presented at the meeting of the Northeastern Educational Research Association, Grossingers, New York, November 1970. (Forster, Passmore)

Levine, M. V., & Rubin, D. B. Measuring the appropriateness of multiple choice tests. Research Bulletin 76-31. Princeton, NJ: Educational Testing Service, 1976.

Lieberman, M. Estimation of a moral judgment level using items whose alternatives form a graded scale. Paper presented at the meeting of the American Educational Research Association, New York, February 1971. The limitations in the methods which are usually used in stage theory test analysis to determine the stage of development of an individual are outlined. Piaget's and Kohlberg's conceptions of moral judgment are presented along with one method of estimating a subject's level of development. The concept of a latent trait is developed. Its relationship to a stage sequence, through stating the probability of being in each stage, is illustrated in a non-technical manner. Using nine situations or stories pointing up a moral dilemma, the stage boundaries within stories, the discriminating powers of the stories, and the latent trait value parameters are derived using maximum likelihood procedures. The calculation of discriminating power requires the derivation of a new form of polychric correlation coefficient. Various studies reported in the literature concerning differences in moral judgment among ages, sexes, socioeconomic classes and culture are replicated and the procedures developed here applied to the data. Finally, the tests of the assumptions required by the model, the results of the above replications, and the generalizability of the method of analysis are discussed in some detail.

Lind, D. Stochastic test models and statistical decision-processes. Zeitschrift fur Experimentelle und Angewandte Psychologie, 1976, 23(3), 446-476.

Lindsey, J. K. Conditional independence and log linear models for multidimensional contingency tables. Quality and Quantity, 1974, 8(4), 337-390.

Linn, R. L., Rock, D. A., & Cleary, T. A. Sequential testing for dichotomous decisions. Educational and Psychological Measurement, 1972, 32, 85-95. (Also Research Bulletin 70-31. Princeton, NJ: Educational Testing Service, 1970).

Lippert, E., Schneider, P., & Wakenhut, R. Stability of probabilistic scaling methods. Psychologische Beitrage, 1977, 19(4), 588-599.

Lippert, E., & Wakenhut, R. Centrality of attitudes. Zeitschrift fur Soziologie, 1978, 7(1), 87-96.

Loevinger, J. Person and population as psychometric concepts. Psychological Review, 1965, 72, 143-155. (Forster, Passmore)

Lord, F. M. A theory of test scores. Psychometric Monograph, 1952, No. 7.

Lord, F. M. The relation of test score to the trait underlying the test. Educational and Psychological Measurement, 1953, 13, 517-548. (Also Research Bulletin 52-10. Princeton, NJ: Educational Testing Service, 1952).

Lord, F. M. A note on the normal ogive or logistic curve theory. Psychometrika, September 1965, 30, 371-372. (Also Research Bulletin 64-56. Princeton, NJ: Educational Testing Service, 1964). (Roberts)

Lord, F. M. An empirical study of item-test regression. Psychometrika, September 1965, 30, 373-376. (Also Research Bulletin 65-21. Princeton, NJ: Educational Testing Service, 1965). (Roberts)

Lord, F. M. An analysis of the Verbal Scholastic Aptitude Test using Birnbaum's three-parameter logistic model. Educational and Psychological Measurement, 1968, 28, 989-1020. (Also Research Bulletin 67-34. Princeton, NJ: Educational Testing Service, 1967)

Lord, F. M. Estimating true-score distributions in psychological testing (An empirical Bayes estimation problem). Psychometrika, 1969, 34, 259-299. (Also Research Bulletin 67-37. Princeton, NJ: Educational Testing Service, 1967)

Lord, F M. Estimating item characteristic curves without knowledge of their mathematical form. Psychometrika, 1970, 35, 43-50. (Also Research Bulletin 68-8. Princeton, NJ: Educational Testing Service, 1968)

Lord, F. M. Some test theory for tailored testing. In W. H. Holtzman (Ed.), Computer-Assisted Instruction, Testing, and Guidance. New York: Harper and Row, 1970. (Also Research Bulletin 68-38. Princeton, NJ: Educational Testing Service, 1968)

Lord, F. M. Robbins-Munro procedures for tailored testing. Educational and Psychological Measurement, 1971, 31;, 3-31. (Also Research Bulletin 69-18. Princeton, NJ: Educational Testing Service, 1969)

Lord, F. M. A theoretical study of two-stage testing. Psychometrika, 1971, 36, 227-242. (Also Research Bulletin 69-95. Princeton, NJ: Educational Testing Service, 1969)

Lord, F. M. Significance test for a partial correlation corrected for attenuation. Research Bulletin 72-6. Princeton, NJ: Educational Testing Service, 1972.

Lord, F. M. Individualized testing and item characteristic curve theory. Research Bulletin 72-50. Princeton, NJ: Educational Testing Service, 1972. (Forster, Passmore)

Lord, F. M. Anchored scaling and equating: Old conceptual problems and new methods. Research Bulletin 72-78. Princeton, NJ: Educational Testing Service, 1972. (Forster, Passmore)

Lord, F. M. Power scores estimated by item characteristic curves. Educational and Psychological Measurement, 1973, 33, 219-224. (Also Research Bulletin 72-46. Princeton, NJ: Educational Testing Service, 1972)

Lord, F. M. Individualized testing and item characteristic curve theory. In D. H. Krantz, R. C. Atkinson, R. D. Luce, & P. Suppes (Eds.), Contemporary developments in mathematical psychology, Vol. II. San Francisco: Freeman, 1974. (Also Research Bulletin 72-50. Princeton, NJ: Educational Testing Service, 1972)

Lord, F. M. Estimation of latent ability and item parameters when there are omitted responses. Psychometrika, 1974, 39, 247-264. (Also Research Bulletin 73-37. Princeton, NJ: Educational Testing Service, 1973)

Lord, F. M. The relative efficiency of two tests as a function of ability level. Psychometrika, 1974, 39, 351-358. (Also Research Bulletin 73-41. Princeton, NJ: Educational Testing Service, 1973)

Lord, F. M. Quick estimates of the relative efficiency of two tests as a function of ability level. Journal of Educational Measurement, 1974, 11, 247-254. (Also Research Bulletin 73-67. Princeton, NJ: Educational Testing Service, 1973)

Lord, F. M. Practical methods for redesigning a homogeneous test, also for designing a multi-level test. Research Bulletin 74-30. Princeton, NJ: Educational Testing Service, 1974.

Lord, F. M. The ability scale in item characteristic curve theory. Psychometrika, June 1975, 40, 205-217. (Also Research Bulletin 74-19. Princeton, NJ: Educational Testing Service, 1974) (Roberts)

Lord, F. M. Relative efficiency of number-right and formula scores. British Journal of Mathematical and Statistical Psychology, 1975, 28, 46-50. (Also Research Bulletin 74-9. Princeton, NJ: Educational Testing Service, 1974.)

Lord, F. M. Evaluation with artificial data of a procedure for estimating ability and item characteristic curve parameters. Research Bulletin 75-33. Princeton, NJ: Educational Testing Service, 1975.

Lord. F. M. A survey of equating methods based on item characteristic curve theory. Research Bulletin 75-13. Princeton, NJ: Educational Testing Service, 1975. (Roberts)

Lord, F. M. A broad range tailored test of verbal ability. Proceedings of the Conference on Computerized Testing. Washington, DC, June, 1975. U. S. Civil Service Commission. (Also Research Bulletin 75-5. Princeton, NJ: Educational Testing Service, 1975.)

Lord, F. M. A study of item bias using item characteristic curve theory. Paper presented at the Third International Association for Cross-Cultural Psychology Congress, Tilburg University, Tilburg, the Netherlands, 1976.

Lord, F. M. Practical applications of item characteristic curve theory. Journal of Educational Measurement, 1977, 14(2), 117-138. (Also Research Bulletin 77-3. Princeton, NJ: Educational Testing Service, 1977.) A variety of practical applications of item characteristic curve test theory are discussed. Among these applications are tailored testing, two stage testing, determining whether two tests measure the same latent trait, and measuring item bias towards minority or other groups.

Lord, F. M., & Novick, M. R. Statistical theories of mental test scores. Reading, MA: Addison-Wesley, 1968.

Loret, P. G., Seder, A., Bianchini, J. C., & Vale, C. A. Anchor Test Study Final Report. Project Report and Volumes 1 through 30. Berkeley, CA: Educational Testing Service, 1974. (ERIC Document Service Nos. 092 601 through ED 092 631).

Lukesch, H., & Tischler, A. Cross-validation of the Marburg Scales for assessing parental educational style. Psychologie und Praxis, 1975, 19(1), 19-30. Attempted to replicate results obtained by K. H. Stapf et al (1972) in studying the Marburg scales. Ss were 206 boys and 217 girls (10-14 yrs. old) from 12 classes in 3 different types of schools in Austria. The following test statistics are reported for the total sample as well as for boys and girls separately: item difficulty and item-test correlations for each of the 60 items: means, standard deviations, reliabilities, and average item-test correlations for each of the 4 subscales (a) using the 5-point response scale and (b) after dichotomizing the answers. Compared to the perception by boys, girls saw their mothers as more supporting and both parents as less firm. Results of 4 factor analyses (separate for boys and girls and for perceptions of fathers and mothers) are described. Two factors, Support and Firmness, explained 26-48% of the variance. The factor structure was similar for boys and girls and resembled the results by Stapf. An attempt to apply the Rasch model to the data failed, and it is concluded that the Marburg scales in their present form are not scalable according to the Rasch model. (Roberts)

Lumsden, J. Test theory. Annual Review of Psychology, 1976, 27, 251-280.

Lumsden, J. Tests are perfectly reliable. British Journal of Mathematical and Statistical Psychology, 1978, 31(5), 84-98.

Mandeville, G. K., & Smarr, A. M. Rasch Model analysis of three types of cognitive tests data. In J. P. Ryan (Chair), Critical perspectives on the Rasch latent trait model. Symposium presented at the meeting of the American Educational Research Association, San Francisco, 1976.

Marco, G. L. Use of the logistic model as an alternative to linear interpolation for computing percentile ranks. Research Memorandum 75-6. Princeton, NJ: Educational Testing Service, October 1975. (Roberts)

Marco, G. L. Item characteristic curve solutions to three intractable testing problems. Journal of Educational Measurement, 1977, 14(2), 138-160. This paper summarizes three studies that illustrate how application of the three-parameter logistic test model helped solve three relatively intractable testing problems. The three problems are: designing a multi-purpose test, evaluating a multi-level test, and equating a test on the basis of pretest statistics.

*Martois, J. S. The organization and implementation of a minimum competency testing program. Paper presented at the 1978 Conference on Minimum Competency Testing, Washington, DC (8 pages). (Harvey)

Maxey, E. J. An investigation of the Rasch probability model for speed in an oral reading test. (Doctoral dissertation, University of Iowa, 1967). Dissertation Abstracts International, 28(8-A), 3075.

McBride, J. R. Bandwidth, fidelity, and adaptive tests. In CATC-2 1975. Atlanta, GA: Atlanta Public Schools, 1976, 81-98.

McBride, J. R., & Weiss, D. J. TETRATEST: A Fortran IV Program for calculating tetrachoric correlatinos. Research Report 72-2. Minneapolis, MN: University of Minnesota, Department of Psychology, 1975. A general purpose computer program for the calculation of a matrix of tetrachoric correlations is described. This program was developed for use in adaptive (and other) testing research for examining the unidimensionality assumption in latent trait theory, in conjunction with available factor analysis programs. Several other potential applications, as well as details for its use, are described. The program accepts as input raw dichotomous data, reduced joint frequency data, or joint and marginal proportions, for up to 75 items. Output options include the tetrachoric correlation matrix, the matrix of phi coefficients, fourfold frequency tables for every item pair, a joint frequency matrix (which reduces all the information in the fourfold tables to a square matrix with order equal to the number of items), and a pair-by-pair listing of input proportions and output correlations which permits testing the program against published tables of the tetrachoric correlation. Variable input and output formatting makes the program convenient to use in conjunction with other analyses by packaged statistical programs. Examples of input and output are presented. A complete FORTRAN listing is included.

McKay, H., Sinisterra, L., McKay A., Gomez, H., & Lloreda, P. Improving cognitive-ability in chronically deprived children. Science, 1978, 200(4339), 270-278.

Mead, R. Assessing the fit of data to the Rasch model. Paper presented at the American Educational Research Association, San Francisco, April 1976. This paper considers (1) the requirements imposed on data in order to conform to the Rasch model, (2) some common sources of departure from the model, and (3) a procedure for recognizing the occurrence of these disturbances. The specific disturbances discussed are guessing, practice, speededness, and bias. The observed characteristic curve for each situation is compared with the true logistic ogive and with the item characteristic curve that the Rasch estimation procedure would fit to such data. A convenient, interpretable form of residual between model and data is suggested.

Mead, R. J. Assessing the fit of data to the Rasch model through analysis of residuals. Doctoral dissertation, University of Chicago, 1976.

Mellenbergh, G. H. An application of a model of Rasch. Unpublished manuscript, University of Amsterdam, The Netherlands, 1969. (Forster, Passmore)

Meredith, W. Poisson distributions of error in mental test theory. British Journal of Mathematical and Statistical Psychology, 1971, 24(5), 49.

Meredith, W., & Kearns, J. Empirical Bayes point estimates of latent trait scores without knowledge of the trait distribution. Psychometrika, 1973, 38(4, Pt. 1), 533-554. Connects recent developments in empirical Bayes procedures with current work in mental test theory Point estimators of true scores are derived for the binomial and Rasch test models. These estimators are shown to be asymptotically optimal. Smoothing and an empirical study of the behavior of empirical Bayes estimates are included. (Roberts)

Meuwese, W. Application of sample-free item analysis on achievement tests Nederlands Tijdschrift voor de psychologie en haar Grensgebiden, 1970, 25, (9), 585-593. (Psychological Abstracts, 1971, 46, No. 3730) (Forster, Passmore)

Micko, H. C. A generalization of the metric model of Rasch and its application to the psychophysics of reactions. Psychologische Beitrage, 1970, 12(1), 4-22. Enlarged the logistic metric model of G. Rasch to N parameters that may be related to situations, persons or reactions. A model of 3 parameters is applied to the measurement of RT. (Roberts)

Milholland, J. E. A note on person-free item calibration. Paper presented at the meeting of the American Educational Research Association, Chicago, 1971. (Forster, Passmore)

Minor, J. J., & Fiske, D. W. Response processes during the description of others. Educational and Psychological Measurement, 1976, 36(4), 829-833. Response processes of undergraduates describing familiar peers of the same sex were investigated. Subjects described persons by responding to items adapted from Jackson's Personality Research Form. Subjects also reported item ambiguity, inappropriateness, etc. Results suggest that similar processes are involved in describing self and others.

Moonan, W. J. Evaluating trainee test performances by a G. Rasch measurement model: a dialogue. Paper presented at the 11th Annual Conference of the Military Testing Association, New York, September 17, 1972. (Forster, Passmore)

Moonan, W. J. Item banking with the Rasch model. Unpublished manuscript, Naval Personnel and Training Research Lab, San Diego, CA, 1972. (Forster, Passmore)

Moonan, W. J. PRIME: Personnel research instrumentalities for measurement and evaluation. Paper presented at the meeting of the Military Testing Association, Fort Benjamin Harrison, Indiana, September 1975. The purposes of the paper were to develop a structure by which assessment models could be uniquely characterized and to develop assessment models and their associated computer programs which could be used for personnel research purposes. To these ends two aspects of assessment, measurement and evaluation, were identified. Their distinction was mainly in the process used to respond to the items of an asessment instrument. The parameters of the assessment models were identified. These were variously combined and operationalized into either measurement or evaluation models. The intention of this process was to make evaluation more measurement-like by using and developing evaluation models, for polychotomous response information, which were patterned after Rasch latent-trait measurement models. Also developed were some special models for use in decision making paradigms and for assessment situations for which an estimation sample does or does not exist. Analysis of these models and their associated data were provided by a system of computer programs called PRIME. In all, five programs were developed, illustrated, and documented. Applications of these models and computer programs to personnel research areas of the Navy Personnel Research and Development Center research program was discussed.

Moonan, W. J. & Covher, M. MAXCO: Maximum likelihood estimations of abilities corresponding to raw scores of tests (FACT Document No. 24). San Diego, CA: Naval Personnal and Training Research Laboratory, December 1972.

Mulaik, S. A. Three multidimensional scaling models having strong distribution assumptions. Submitted for publication, 1972. (Forster, Passmore)

Mulaik, S. A. A mathematical investigation of some multidimensional Rasch models for psychological tests. Paper presented at the meeting of the Psychometric Society, Princeton, NJ, March, 1972. (Forster, Passmore)

Murray, D. Rasch item analysis and scaling. Bulletin of the British Psychological Society, 1975, 28(5), 246.

Mussio, J. J. A modification to Lord's model for tailored tests. Unpublished doctoral dissertation, University of Toronto, 1973.

Noack, H. R. Application of latent trait models to the ACT Mathematics Usage Test. (Doctoral dissertation, Iowa State University, 1973). Dissertation Abstracts International, 34(50B), 2288.

Olson, L. A. Postdoctoral fellowship program in educational research, final report. Berkeley, CA: University of California. Sponsoring Agency: Office of Education (DHEW), Washington, D.C., Bureau of Research. During one year as a 1968-69 U. S. Office of Education Postdoctoral Fellow at the University of California at Berkeley, the author

attended three USOE orientation sessions and ten professional conferences and wrote three papers for presentation at these conferences. He conducted research on educational testing and methods of test analysis, and audited several graduate courses in statistics. Reading activity was concentrated on general education in American universities, statistical analysis, and the Rasch method of test analysis. The year's experience led to the development of several ideas for further defining the as yet ambiguous role of a postdoctoral fellow within the host institution. Establishing informal contacts with faculty members, auditing graduate classes, working closely with one particular faculty member, associating oneself with a research project, or teaching for one semester are some of the possible approaches toward optimizing the benefits of a postdoctoral experience.

O'Reilly, R. P., & Others. The validation and refinement of measures of literal comprehension in reading for use in policy research and classroom management. Albany, NY: New York State Education Department, Division of Research, 1976. The report proposes to complete the validation and refinement of a new domain referenced testing technology designed to assess literal comprehension ability in students in grades 1-12. The domain referenced measures in this technology, along with other more traditional measures of reading comprehension, literal and non-literal, are subsequently intended to be used in part in large scale studies of productivity in reading instruction have had little influence on educational decision-making due to serious methodological problems, one of the major problems being the lack of adequate measures of program output. The report further proposes to solve a number of important instructional management problems created by the use of the inadequate information available from traditional measures of reading comprehension. The new domain referenced measures of reading comprehension will have an improved basis for scaling students on comprehension ability, and bility scores from this scale will be referenced to an additional scale defining an individual or group's ability to read in several domains of written discourse. These scaling features will allow for the assignment of students to specific levels of reading materials in specific instructional or content domains, a procedure not possible with existing measures of reading comprehension.

Owen, R. A Bayesian sequential procedure for quantal response in the context of adaptive mental testing. Journal of the American Statistical Association, 1975, 70, 351-356.

Panchapakesan, N. The simple logistic model and mental measurement. Doctoral dissertation, University of Chicago, 1969. (Available from University of Chicago Library). (Forster, Passmore)

Passmore, D. L. Objective measurement in occupational education. Journal of Industrial Teacher Education, (Date unknown) (Forster, Passmore)

Passmore, D. L. A study of the usefulness of weighting test item responses. Minneapolis, MN: Minnesota Research Coordinating Unit in Occupational Education, 1972. (ERIC Document Reproduction Service No. ED 069 692). The purpose of this study was to investigate the practicality of multiple discriminant fuction analysis for deriving item response weights. Item responses on a job satisfaction questionnaire administered to 219 professional workers and 242 semi-skilled customer workers were analyzed. Discriminant functional analysis was conducted on the total sample. Respondents were then randomly assigned to one of two subsamples. Two different discriminant function analyses were then undertaken to maximize group differences in each of the samples. Inconsistency of the results indicates that weights so derived are not generalizable to an independent sample from the same population. Though an increment in predictive efficiency of 11% might be realized with the differential weighting system, the veracity of the increment is doubtful due to the failure of the differential weighting system to cross-validate. Investigating other weighting techniques, such as latent trait measurement models, empirical techniques for weighting each multiple-choice alternative of a test item, and confidence weighting, is suggested.

Passmore, D. L. Objective measurement in occupational education. (ERIC Document Reproduction Service No. ED 069 691), 1972. The model for educational measurement developed by George Rasch, a Danish psychometrician, is reviewed and its application to occupational educational testing is discussed. The Rasch model is an adaptation from the theory of latent trait analysis. According to it, answering an item correctly is a function of the difficulty of the item and the ability of the person being tested. The raw scores serve as the basis for estimating the scale of ability. The author concludes that the properties of the Rasch analysis suggest solutions to a number of measurement problems in occupational engineering including developing and equating alternate forms of a test and estimating and interpreting changes in trainee performance. The item-free characteristics of this measurement model may allow the development of individually tailored tests.

Passmore, D. L. References to the Rasch one parameter logistic measurement model. (ERIC Document Reproduction Service No. ED 078 018), 1973. This bibliogaphy is a list of theoretical and empirical studies of the Rasch One Parameter Logistic Measurement Model. A total of 96 entries are included.

Passmore, D. L. Objective measurement in occupational education. Journal of Industrial Teacher Education, 1973, 10(4), 15-21.

Passmore, D. L. An application of the Rasch one parameter logistic measurement model to the National League for Nursing Achievement Test in Normal Nutrition (Doctoral dissertation, University of Minnesota, 1974). Dissertation Abstracts International, 35(02-A), 963. (University Microfilms No. 74-17271)

Passmore, D. L. Theory and applications of Rasch measurement models: A bibliography. Catalog of Selected Documents in Psychology, 1976, 6,(4).

Patience. W. M. Description of components in tailored testing. Behavior Research Methods and Instrumentation, 1977, 9(2), 153-157.

Pearson, A. A. Empirical results using the Rasch item analysis model. Dissertation Abstracts International, 1974, 34(10-A) 6458.

Perlain, R. The Rasch model as conjoint measurement. (Unpublished paper). University of Chicago, 1977.

Petermann, F., & Knopf, M. Problems in measuring attitude changes .2. New developments and concepts. Zeitschrift fur Sozialpsychologie, 1976, 7(3), 217-230.

Petersen, E. Conception of criminality illustrated by a stochastic process model for deviant behavior. Journal of Research in Crime and Delinquincy, 1972, 9(1), 31-45.

Pine, S. M. Applications of item response theory to the problem of test bias. Unpublished manuscript. Minneapolis: University of Minnesota, Department of Psychology, 1976.

Pine, S. M., & Weiss, D. J. Effects of item characteristics on test fairness. Research Report 76-5. Minneapolis: University of Minnesota, Department of Psychology, 1976.

Poortinga, Y. H. Limitations on intercultural comparison of psychological data. Nederlands Tijdschrift voor de Psychologie en haar Grensgebieden, 1975, 30(1), 23-39. Discusses the comparability of results of testing among diverse cultural groups. A test is acceptable for measuring different populations if it measures qualitatively and quantitatively the same attribute, although test scores of a group are not comparable with scores of other groups if a single relevant condition is not met. A method of determining functional equivalence of tests is the correlational technique, although this procedure may incorporate irrelevant skills. The application of variance is useful when the effects of the main factors and of the interactions can be separated. The item analysis method can reveal latent trait models which can indicate the relationships between distributions of observed item scores and of the underlying traits. It is emphasized that the investigator should not avoid the study of apparently unimportant variations and not seek to find absolute, instead of relative, differences among groups.

Preiser, S. Measurement of stimulus and object features by inversion of psychodiagnostic principles - object-diagnostics as a basis of a differential psychology of environmental conditions. Archiv fur Psychologie, 1977, 129(1), 65-82.

Rasch, G. Probabilistic models for some intelligence and attainment tests. Copenhagen, Denmark: Danmarks Paedogogiske Institut. Nielsen & Lydiche, 1960. (English translation) (Forster, Passmore)

Rasch, G. On general laws and the meaning of measurement in psychology. Proceedings of the fourth Berkeley symposium on mathematical statiscics, 1961, 4, 321-334. (Forster, Passmore)

Rasch, G. Objective comparisons, UNESCO Seminar, Voksenasen, Oslo, 1964. (Forster, Passmore)

Rasch, G. An individualistic approach to item analysis. In P. F. Lazarsfeld and N. W. Henry (Eds.), Readings in Mathematical Social Science. Chicago: Science Research Associates, 1966. (Forster, Passmore)

Rasch, G. An item analysis which takes individual differences into account. British Journal of Mathematical and Statistical Psychology, 1966, 19(1) 49-57. (Forster, Passmore, Roberts)

Rasch, G. An informal report of objectivity in comparisons. In L. J. Van Der Kamp & C. A. J. Viek (Eds.) Psychological measurement theory, Proceedings of the NUFFIC international summer session in science at "HetOude Hof," Den Haag, July 14-28, 1966. Leiden: 1967. (Forster, Passmore)

Rasch, G. A mathematical theory of objectivity and its consequences for model construction. Report from European Meeting on Statistics, Econometrics and Management Sciences, Amsterdam, 1968. (Forster, Passmore)

Rasch, G. Lectures at the University of Georgia, Athens, Georgia, March, 1973.

Rasch, G. Objectivity in psychometrics. A paper presented at the meeting of the Psychometric Society, Chicago, 1973. (Forster, Passmore)

Reckase, M. D. Development and application of a multivariate logistic latent trait model. (Doctoral dissertation, Syracuse University, 1972). Dissertation Abstracts International, 33(9-B), 4495.

Reckase, M. D. An application of the Rasch simple logistic model to tailored testing. Paper presented at the meeting of the American Educational Research Association, Chicago 1974. An application of the two-parameter logistic model to tailored testing is presented. The model is discussed along with the maximum likelihood estimation of the ability parameters given the response pattern and easiness parameter estimates for the items. The technique has been programmed for use with an interactive computer terminal. Use of the procedure is described in a flexible achievement testing setting. Results are presented showing the number of items needed for good estimation. The independence of items used and ability estimation is shown. Applications of the system to intelligence testing are discussed.

Reckase, M. D. Interactive computer program for tailored testing based on one-parameter logistic model. Behavior Research Methods and Instrumentation, 1974, 6(2), 208-212.

Reckase, M. D. The effect of item choice on ability estimation when using a simple logistic tailored testing model. Paper presented at the meeting of the American Educational Research Association, Washington, D. C., 1975. This paper explores the effects of item choice on ability estimation when using a tailored testing procedure based on the Rasch simple logistic model. Most studies of the simple logistic model imply that ability estimates are totally independent of the items used, regardless of the testing procedure. This paper shows that the ability estimate is affected by item choice and gives theoretical and empirical information as to the direction and magnitude of the effect.

Reckase, M. D. Procedures for computerized testing. Behavior Research Methods and Instrumentation, 1977, 9(2), 148-152.

Reckase, M. D. Ability estimation and item calibration using the one and three parameter logistic models: A comparative study. Research Report 77-1. Columbia, MO: University of Missouri, Tailored Testing Research Lab., 1977. Latent trait model calibration procedures were used on data obtained from a group testing program. The one-parameter model of Wright and Panchapakesan and the three-parameter logistic model of Wingersky, Wood, and Lord were selected for comparison. These models and their corresponding estimation procedures were compared, using actual and simulated test results, with respect to: (1) the ability to calibrate multivariate data, (2) the sample size needed for calibration, (3) the effects of item quality on the calibration, and (4) the operational costs. The one-parameter procedure provided equivalent ability estimates at a lower cost than the three-parameter procedure when basically unifactor group tests were used. This method was therefore recommended for item calibration of group administered, multiple choice tests. However, the goodness of fit of the models to the data definitely showed the three-parameter model to be superior. Whether the better fit will outweigh the higher costs can only be determined by further comparison of the usefulness of the ability estimates in actual tailored testing.

Reckase, M. D. Computerized achievement testing using the simple logistic model. Paper presented at the meeting of the American Educational Research Association, New York, April 1977. The reliability and validity of a tailored testing procedure based on the simple logistic model was determined for an achievement test in statistics and measurement. The test was administered on a CRT terminal to students from graduate and undergraduate measurement courses. Equivalent form reliability over a one-week interval was found to be 0.595 while KR20 reliability measures for the traditional course exams yielded 0.756 and 0.574. The tailored testing procedure administered 20 items or less while the traditional tests were 55 and 50 items respectively. The tailored test was a valid predictor for the graduate course.

Reckase, M. D. A comparion of the one- and three-paramerter logistic models for item calibration. Paper presented at the meeting of the American Educational Research Association, Ontario, Canada, March 1978. Five comparisons were made relative to the quality of estimates of ability parameters and item calibrations obtained from the one-parameter and three-parameter logistic models. The results indicate: (1) the three-parameter model fit the test data better in all cases than did the one-parameter model. For simulation data sets, multi-factor data were less well fit than single-factor data. (2) The one-parameter model ability estimates shared more variance with the item responses than did the three-parameter model. (3) There was no difference in the concurrent validity for small samples between the two models in predicting classroom achievement tests. (4) The three-parameter model required larger samples for calibration than did the one-parameter model. (5) The ability estimates from the two models correlated highly for most of the data sets. The one-parameter model is preferred for use with small sample data; but the goodness of fit data parameters is important. The three-parameter model fit all data sets better than the one-parameter model. Data sets from the Missouri School and College Ability Tests, and from undergraduate course final examinations were used to illustrate the models.

Rentz, C. C. An investigation of the invariance properties of the Rasch Model parameter. (Doctoral dissertation, University of Georgia, 1975) Dissertation Abstract International, 36(08-B), 4139. (University Microfilms No. 76-02258)

Rentz, C. C. Rasch model invariance as a function of the shape of the sample distribution and degree of model-data fit. Paper presented at the meeting of the Florida Educational Research Association, Orlando, Florida, January 1976.

Rentz, R. R. Applications of the Rasch Model: Test equating and interpretation. Invited address at the Conference on the Rasch-Wright Model, Indiana University, Bloomington, Indiana, November 1971. (Forster, Passmore)

Rentz, R. R. Technical manual: Language skills examination. Atlanta, GA: University System of Georgia, Board of Regents, 197.. (Forster, Passmore)

*Rentz, R. R. Monitoring the quality of an item-pool calibrated by the Rasch Model. Paper presented at the meeting of the National Council on Measurement in Education, Toronto, Ontario, March, 1978.

Rentz, R. R. & Bashaw, W. L. Evaluation of the Rasch strong true score model as a method for equating non-parallel reading achievement tests. Research proposal presented to the U. S. Office of Education, 1972. (Forster, Passmore)

Rentz, R. R., & Bashaw, W. L. Equating reading tests with the Rasch model, volume I, final report Athens, GA: University of Georgia, Educational Research Laboratory, 1975. (ERIC Document Reproduction Service No. ED 127 330). In order to determine if Rasch model procedures have any utility for equating pre-existing tests, this study reanalyzed the data from the equating phase of the Anchor Test Study which used a variety of equipercentile and linear model methods. The tests involved included seven reading test batteries, each having from one to three levels and two forms, and each having a vocabulary and comprehension subtest. There were 28 form-level combinations possible. Therefore, of concern was the simultaneous equating of 28 tests for each of vocabulary, comprehehsion and total scores. Seven objectives characterized the study and are elaborated on in separate sections of this report. The objectives were to: (1) describe a methodology for test equating using the Rasch model, (2) describe basic item analysis data for each test in the Anchor Test Study data base, (3) evaluate the fit of the Rasch model with respect to those tests that were part of the data base, (4) investigate the stability of Rasch model parameter estimates under conditions of varying sample size and sample composition, (5) provide tables of equating based on Rasch model methods, (6) estimate the equating error associated with the use of these equating methods, and (7) compare the results of equating with those obtained in the Anchor Test Study.

Rentz, R. R., & Bashaw, W. L. Equating reading tests with the Rasch model, volume II, technical reference tables. Athens, GA: University of Georgia, Educational Research Laboratory, 1975. (ERIC Document Reproduction Service No. ED 127 331). This volume contains tables of item analysis results obtained by following procedures associated with the Rasch model for those reading tests used in the Anchor Test Study. Appendix I gives the test names and their corresponding analysis code numbers. Section I (Basic Item Analyses) presents data for the item analysis of each test in a two part table. The first part, Item Summary, gives information for each item in the test. The second part, Ability Summary, shows the frequency of distribution and the calibrated Rasch scores and their asociated errors of measurement. At the top of each table appears the following information: number of questions, the adjusted sample size, the traditional Kuder-Richardson formula 20 reliability estimate, and the equating constant. Section II presents summaries of traditional and Rasch item parameter estimates of each test. This section shows the distribution of various item statistics and the relationship between selected statistics. Relationships are described graphically. Each table gives the test name and analysis code number, the number of items, and the sample size on which the page of data is based. The first table gives the usual summary statistics. The second table, Tally by Classification, presents the frequency distributions of the item statistics.

Rentz, R. R., & Bashaw, W. L. Equating reading tests with the Rasch model. Management report. Athens, GA: University of Georgia, Educational Research Laboratory, 1976. (ERIC Document Reproduction Service No. ED 127 329). To determine if Rasch model procedures have any utility for equating pre-existing tests, data from the equating phase of the Anchor Test Study (ATS) were reanalyzed. This management report summarizes the work completed in the project, describes the differences in this project and that of the ATS, and presents recommendations and conclusions to the U. S. Office of Education. The tests involved included seven reading test batteries, each having one to three levels and two forms, and each having a vocabulary and comprehension subtest. There were 28 form-level combinations possible. Therefore, the concern was the simultaneous equating of 28 tests for each of vocabulary, comprehension, and total scores. The obvious difference between this project and the ATS is that this project used Rasch model test calibrations and equating methods while the ATS used a variety of equipercentile and linear model methods. Other important differences are outlined. Finally, some conclusions and recommendations relative to any future equating effort that might be undertaken are made. Topics discussed include technical issues as well as design and cost considerations.

Rentz, R. R. & Bashaw, W. L. The national reference scale for reading: An application of the Rasch Model. Journal of Educational Measurement, 1977, 14, 161-179. Describes the characteristics, properties, and development of the National Reference Scale (NRS) for reading. This new scale is the result of a reanalysis of the Anchor Test Study data using Rasch model procedures to produce equated scores among all of the reading tests included in that study. The NRS is a scale that spans 28 reading tests: it enables the test user to utilize those tests interchangeably. An evaluation of the appropriateness of the Rasch model for equating these tests is provided, as well as a general theoretical framework for evaluating fit of the Rasch model. (Roberts)

Rentz, R. R., Bashaw, W. L., Cartledge, C. G., & Brigman, S. L. Equating reading tests with the Rasch model (2 vols.). Athens, GA: Educational Research Laboratory, 1975.

Rentz, R. R. & Rentz, C. C. Does the Rasch Model really work? A discussion for practitioners. Prepublication draft, 1978.

*Rentz, R. R., & Ridenour, S. E. The fit of the Rasch Model to achievement tests. Paper presented at the meeting of the Eastern Educational Research Association, Williamsburg, VA, March 1978. (Roberts)

Repp, B. H. Measuring laterality effects in dochotic-listening. Journal of the Acoustical Society of America, 1977, 62(3), 720-737.

Ross, J. An empirical study of a logistic mental test model. Psychometrika, 1966, 31, 325-340. (Also Research Bulletin 65-11. Princeton, NJ: Educational Testing Service, 1965)

Ross, J., & Lumsden, J. Attribute and reliability. British Journal of Mathematical and Statistical Psychology, 1968, 21, 251-263. An analysis of classical test theory and of more recent developments shows that the concepts of true score, parallel test, and reliability are poorly defined. Where reliability is defined as generalizability, it is not so much poorly defined as detached from the essential interior properties of the test. It is proposed that tests as such be restrictively defined as devices which map an attribute in a single dimension. A logistic model which fits this conception is examined, and it is shown that true score, parallel test, and reliability can be defined without ambiguity, inconsistency, or circularity. Deeper investigation reveals that the true score concept is unnecessary as a basis of reliability and a distraction. Reliability is formulated as a characterization of the test's mapping. It is shown that the mapping can be characterized for the model and that there is a distinction between information loss due to regression and error which probably lies behind the poorly conceived notion of true score.

Rudner, L. M. A closer look at latent trait parameter invariance. Paper presented at the meeting of the New England Educational Research Organization, Manchester, NH, May 1977. Using conventional mental test theory, item parameters of an aptitude or achievement test vary with each group of examinees, and as such are somewhat limited in their use and interpretation. Within the last 25 years, measurement models have emerged in which item parameters are considered to be invariant. Generically referred to as latent trait mental measurement or item characteristic curve theory, these models have precipitated the development of sophisticated tailored testing testing techniques which capitalize on the item parameter invariance feature. This paper examines the invariance feature for the Birnbaum 2 and 3 parameter logistic models. The investigation is in two parts, the first theoretical, the second empirical. Even with somewhat inadequate data, the empirical investigation supported the theoretical considerations.

Rudner, L. M. Efforts toward the development of unbiased selection and assessment instruments. Paper presented at the International Symposium on Educational Testing, University of Leyden, The Netherlands, June 1977. Investigations into item bias provide an empirical basis for the identification of test items which appear to measure different traits across populations or cultural groups. The psychometric rationales for six approaches to the identification of biased test items are reviewed: (1) Transformed item difficulties: within-group p-values are standardized and compared between groups. (2) Analysis of variance: bias is operationally defined in terms of significant item by group interaction effects. (3) Chi-square: individual items are investigated in terms of between group score level differences in expected and observed proportions of correct responses. (4) Item characteristic curve theory: differences in the probabilities of a correct response, given examinees of the same underlying ability and different culture groups, are

evaluated. (5) Factor analytic: Item bias is investigated in terms of culture specific and culture common sources of variance, or in terms of loadings on a biased test factor. (6) Distractor response analysis: the relative attractiveness of item foils, or response sets, is investigated. The limitations and advantages of each approach in terms of the underlying assumptions, psychometric soundness, conceptual complexity, applicability to criterion referenced testa and applicability to interdependent groups are discussed.

Rudner, L. M. An approach to biased item identification using latent trait measurement theory. Paper presented at the meeting of the American Educational Research Association, New York, April 1977. Because it is a true score model employing item parameters which are independent of the examined sample, item characteristic curve theory (ICC) offers several advantages over classical measurement theory. In this paper an approach to biased item indentification using ICC theory is described and applied. The ICC theory approach is attractive in that it, (1) appears to be sensitive largely to cultural variations in the trait gauged by test items, (2) does not assume total scores to be valid indicators of true ability, (3) places the identified degree of item bias on a quantified metric, and (4) is applicable to items of sufficiently varying degrees of difficulty. While sensitive to some factors other than item bias, namely, local independence, item inappropriateness and poor parameter estimates, the approach may prove useful to the measurement field.

Ryan, A. S. The development and empirical evaluation of a partial-knowledge scoring scheme based on latent trait theory and multiple-category response weights. (Doctoral dissertation, Florida State University, 1974). Dissertation Abstracts International, 35(9-A), 5694.

Ryan, J. P. Comparing classical and latent trait analyses of attitudinal data. Paper presented at the Southeastern Invitational Conference on Measurement in Education, Virginia Polytechnic Institute and State University, December 1977.

Ryan, J. P., & Hamm, D. W. Practical procedures for increasing the reliability of classroom tests by using the Rasch model. Paper presented at the meeting of the National Council on Measurement in Education, San Francisco, April 1976. A procedure is described for increasing the reliability of tests after they have been given and for developing shorter but more reliable tests. Eight tests administered to 200 graduate students studying educational research are analyzed. The analysis considers the original tests, the items loading on the first factor of the test, and the items which fit the Rasch model. Adjustments for test length are considered. Tests shortened by deleting items which do not fit the Rasch model have a higher internal consistency reliability than the longer original tests. This finding contradicts the theorem of classical test theory which states that increasing test length increases test reliability. The explanation for the find is that the theorem of classical test theory assumes that all items are homogeneous. Items not fitting the Rasch model are not unidimensional in the "latent-trait" sense which

corresponds to being non-homogeneous in the "classical-test-theory" sense. Deleting such items results in a proportionately longer test in that a larger proportion of the remaining items is homogeneous. The misfitting items would not have been deleted on the basis of a classical item analysis.

Ryan, J. P, & Hamm, D. W. An introduction to the Rasch latent trait psychometric model. Columbia, SC: University of South Carolina, College of Education; Richland School District #1, 1978

Samejima, F. Estimating latent ability using a pattern of graded scores. Psychometrika Monograph Supplement, 1969, No. 17.

Samejima, F. A general model for free-response data. Psychometric Monograph, 1972, No. 18. Proposes a model based on the latent trait model which can be used to measure a specified psychological behavior or process. The latent trait model is expanded to include items not scored. Situations in which the scoring of items varies are systematized. General conditions for the operating characteristic of an item response category are examined in order to obtain a unique likelihood estimate. It is suggested that by using this model diverse and general psychological constructs, e.g., perceptual sensitivity, mental ability, personality traits, and attitudes may be studied.

Samejima, F. A paradox in the knowlege of guessing model for the multiple-choice item. Proceedings of the Annual Convention of the American Psychological Association, 1972, 7(Pt. 1), 61-62. Following the knowledge or guessing model for the multiple-choice item in the latent trait theory, observations are made that, by the joint scoring, the conditional test score distribution of "pure guessers" can be approached to the 1-point distribution, which may be considered as a merit of the joint scoring; at the same time, the joint scoring brings information loss especially for a higher range of ability. The joint scoring may have advantages for discriminating examinees of lower levels of ability, and the use of both the joint scoring and the simple scoring for 1 set of data may be suggestive.

Samejima, F. Homogeneous case of the continuous response mode. Psychometrika, 1973, 38(2), 203-219. In line with the latent trait model, the continuous response level is defined and considered, in contrast to the discrete response levels, which have already been explored by the author.

Samejima, F. A comment on Birnbaum's three-parameter logistic model in the latent trait theory. Psychometrika, 1973, 38(2), 221-223. Considers A. Birnbaum's 1968 model for the multiple-choice item in the latent trait theory with respect to the item response information function and the unique maximum condition. It is clarified that with models of knowledge or random guessing nature, which include the 3- parameter logistic model, the unique maximum condition is not satisfied for the correct answer, and the item response information function is negative for the interval - , qg. It is suggested that qg should be used as a criterion in selecting optimal items for a specified group of examinees, so that the possibility of nonunique maxima of the likelihood function on the response pattern given by an examinee in the group can be practically avoided.

Samejima, F. Normal ogive model on the continuous response level in the multidimensional latent space. Psychometrika, 1974, 39, 111-121.

Samejima, F. Estimation of the operating characteristics of item response categories I: Introduction to the two-parameter beta method. Research Report 77-1. Knoxville, TN: Tennessee Univerity, Department of Psychology, 1977. A method of estimating the operating characteristics of a new test item added to an established test without assuming a prior model has been proposed by the author. In this paper, the author extends this logic by proposing a different hypothesized distribution of the resulting estimate. The former method used a normal approximation; the present one uses a beta distribution. Both methods fall under the general principles of latent trait test theory.

Samejima, F. Weakly parallel tests in latent trait theory with some criticisms of classical test theory. Psychometrika, 1977, 42(2), 193-198. A new concept of weakly parallel tests, in contrast to strongly parallel tests in latent trait theory, is proposed. Some criticisms of the fundamental concepts in classical test theory, such as the reliability of a test and the standard error of estimation, are given.

Samejima, F. A method of estimating item characteristic functions using the maximum likelihood estimate of ability. Psychometrika, 1977, 42(2), 163-191. A method of estimating item characteristic functions is proposed, in which a set of test items, whose operating characteristics are known and which give a constant test information function for a wide range of ability, are used. The method is based on maximum likelihood estimation procedures.

Sanathan, L. Models and estimation methods in visual scanning experiments. Technometrics, 1972, 14(4), 813.

Sanathan, L. Some properties of logistic model for dichotomous response. Journal of the American Statistical Association, 1974, 69(347), 744-749.

Scheiblechner, H. H. Conditional maximum likelihood parameter estimation in a generalized multifactorial version of Rasch's probabilistic measurement model with two categories of answers. Research Bulletin No. 4. Vienna, Austria: The University of Vienna, Psychological Institut, 1971. A psychological theory with the claim to generality must be independent of incidental factors contributed by the sample upon which the theory is based. A class of measuring models developed by G. Rasch fulfills these demands. Since predicative logic is involved in most reasoning processes it was chosen as topic for the experimental collection of data to develop a theory of reasoning. In order to avoid the symbolism of formal logic, the basic theorems were transformed into simple graphs. Tests with 470 pupils, grades 3 through 8, demonstrated that a single ability can handle all logical operations. Probably only a smaller number of operations than those contained in the operations of predicate logic is employed by Ss. The learning process of such problems may be called component specific, because isolated and independent exercises were learned 1st not a complete composite solution. Importance of this finding for strategies in education is discussed. (Forster, Passmore, Roberts)

Scheiblechner, H. H. Separation of individual- and system-influences on behavior in social contexts. Acta Psychologica, 1971, 35(6), 442.

Scheiblechner, H. H. Learning and solution of complex reasoning problems: A theoretical investigation concerning the complexity of composite predicate logical tasks in tests. Zeitschrift fur Experimentelle und Angewandte Psychologie, 1972, 19(3), 476-506. A psychological theory with the claim to generality must be independent of incidental factors contributed by the sample upon which the theory is based. A class of measuring models developed by G. Rasch fulfills these demands. Since predicative logic is involved in most reasoning processes it was chosen as topic for the experimental collection of data to develop a theory of reasoning. In order to avoid the symbolism of formal logic, the basic theorems were transformed into simple graphs. Tests with 470 pupils, grades 3 through 8, demonstrated that a single ability can handle all logical operations. Probably only a smaller number of operations than those contained in the operations of predicate logic is employed by Ss. The learning process of such problems may be called component specific, because isolated and independent exercises were learned 1st not a complete composite solution. Importance of this finding for strategies in education is discussed.

Schildkamp-Kundiger. Advantages and possibilities of application of linear logistic test models in the area of social sciences. Theory and Decision, 1976, 7(1-2), 107-118. Briefly reviews the assumptions of linear logistic models as developed by G. Rasch (1960) and considers the advantages of such models compared with models of classical test theory. If a satisfactory fit between model and data is obtained, the logistic model will provide information about the dimensionality of latent variables and sample-free estimates of person and item parameters. Examples of the application of logistic models to cognitive and social psychology and to the study of effects of mass media are discussed. (Roberts)

Schmidt, F. L. The Urry method of approximating the item parameters of latent trait theory. Educational and Psychological Measurement, 1977, 37(3), 613-620. Urry's procedure for approximating latent trait test models is shown to tend to underestimate item discriminatory power and overestimate item difficulty. A method for correcting these biases is provided, and implications of the procedures are discussed.

Sitgreaves, R. Review of G. Rasch, probabilistic models for some intelligence and attainment tests. Psychometrika, 1963, 28, 219-220. (Forster, Passmore)

Sixtl, F. Novel basis for logistic function. Zeitschrift for Experimentelle und Angewandte Psychologie, 1973, 20(3), 489-497.

Sixtl, F. Probabilistic unfolding. Psychometrika and Supplements, 1973, 28(2), 235-248.

Slinde, J. A., & Linn, R. L. Vertically equated tests: fact or phantom? Journal of Educational Measurement, 1977, 14, 23-32.

Slinde, J. A., & Linn, R. L. An exploration of the adequacy of the Rasch Model for the problem of vertical equating. Journal of Educational Measurement, 1978, 15, 23-25. (Roberts)

Smith, H. A., & Kifer, E. Concepts, models and methodologies for evaluation of experimental education in pharmacy. American Journal of Pharmaceutical Education, 1978, 42(2), 159-166.

Soriyan, M. A. Measurement of the goodness-of-fit of Rasch's probabilistic model of item analysis to objective achievement test of the West African Certificate Examination. (Doctoral dissertation, University of Pittsburgh). Ann Arbor, MI: University Microfilms, No. 72-7895, 1971. (Forster, Passmore)

Spanda, H. Measurement as an element of curricular evaluation. Studies in Educational Evaluation, 1976, 2(2), 85-102. A properly chosen test model defines the kind of dependence of the observable behavior on underlying abilities of the persons tested in the form of a data-adequate theory. The limitations and potential of several test models are detailed; and they are critically compared as an aspect of empirical evaluation.

Sporli, S. Dedifferentiation as a field effect in the road traffic system: Toward a system theory of the person. Psychologie: Schweizerische Zeitschrift for Psychologie und ihre Andwendungen, 1974, 33(4), 384-406. Questions the usual emphasis on interindividual "latent trait" variance in differential psychology and psychodiagnostics. Instead, it is suggested that road traffic should be understood as a system leading to behavior deformations that are essentially field-specific and only partially person-specific. Possible causes and mechanisms of these deformations are discussed, and possible ways for verifying the theory are suggested. The problem is also considered in a metatheoretical perspective.

Sprung, L. Measurement of sociopolitical attitudes using Rasch scale. Zeitschrift fur Psychologie, 1975, 183(4), 424.

Stapf, K. H. Practical experiences with Rasch's metric model in the construction of specific attitude scales. Psychologische Beitrage, 1970, 12(1), 105-114. Reports the construction of scales for determining parental behavior with G. Rasch's logistic test model. The advantages of scaling frequency and intensity previous to the essential application of the model are discussed. As an example a 9-item scale concerning punishment behavior is described. The strict criteria of the model are sufficiently valid for these items.

Stelzl, I. The usefulness of the Rasch model. Psychologische Beitrage, 1972, 14(2), 298-310. Discusses the use of G. Rasch's model for test scores in test construction. The implications of the model for reliability, validity, and standardization of tests are examined. Maximum likelihood confidence intervals for test scores are derived from the model. It is also shown that the common practice of standardization on the basis of the normal distribution does not change the favorable features of the Rasch model scales. (Roberts)

Stiles, R., & Others. Use of the Rasch model with content referenced tests. Paper presented at the meeting of the American Educational Research Association, Washington, D. C., March 1975. The purpose of this study was to explore the usefulness of the Rasch test analysis model for providing consistent scale statistics between content-referenced and norm-referenced tests across school districts. A number of content-referenced tests were created to assess student performance relative to basic skills goals identified in reading and mathematics. Performance on these tests was related to performance on norm-referenced tests in two school districts. Based on preliminary analyses, it appears that the Rasch model offers promise for a fully flexible content-referenced test system providing information now available only through the use of norm-referenced tests.

Swineford, F., & Fan, C. T. A method of score conversion through item statistics. Psychometrika, 1957, 22, 185-188.

Taylor, C. W. Maximizing predictive efficiency for a fixed total testing time. Psychometrika, 150, 15, 391-406.

Thissen, D. M. Incorporating item response latencies in latent trait estimation. Dissertation Abstracts International, 37(9-B), 4658.

Thissen, D. M. Information in wrong responses to Raven progressive matrices. Journal of Educational Measurement, 1976, 13, 201-214.

Thissen, D. M. Response latency and individual differences: A investigation in a general latent trait model. Paper presented at the meeting of the American Educational Research Association, New York, April 1977. Few phenomena in psychology have been studied more than reaction time. However, while response latency now provides the basic data for research in many areas of cognitive psychology, little work has been done on the problem of incorporating response latency in the measurement of individual differences. A solution to this problem requires the construction of a mathematical model relating reaction time to individual and test item variation. The present research is concerned with the development and empirical validation of such a model. The model developed here provides a simultaneous analysis of test item responses and the latencies associated with these responses. The method is sufficiently general to be used in the analysis of a wide variety of data, ranging from responses to conventional ability tests to data collected in experimental problem-solving situations. The application of this model to a set of spatial visualizing ability test data is discussed. The results indicate that the model provides a parsimonious summary of the raw data, and estimates of individual differences useful in further analysis.

Thomson, J. D., & Parkin, B. How do you prefer your item analyses - traditional or Rasch. Australian Journal of Education, 1976, 20(1), 107-109.

Tinsley, H. E. A comparison of the Rasch model with traditional test construction techniques. Unpublished research proposal, University of Minnesota, 1970. (Available from Minnesota Research Coordinating Unit in Vocational Education). (Forster, Passmore)

Tinsley, H. E. An investigation of the Rasch simple logistic model for tests of intelligence or attainment (Doctoral dissertation University of Minnesota, 1971). Dissertation Abstracts International, 1972, 32 11-8), 6629. (University Microfilms No. 72-14387) (Forster, Passmore).

Tinsley, H. E., & Dawis, R. V. A comparison of the Rasch item probability with three common item characteristics as criteria for item selection (Technical Report No 3003). Project No. NR151-323, Personnel and Training Research Programs, Office of Naval Research, 1972. (ERIC Document Reproduction Service No. ED 068 516). Compared the scores of college students on analogy tests differing in item-selection criteria. Results show that G. Rasch's item probability model was a more lenient criteria for selection than item discrimination, item difficulty, and item-ability correlation. Only moderate amount of overlap was found among the 4 criteria. (Forster, Passmore, Roberts)

Tinsley, H. E., & Dawis, R. V. Test-free person measurement with the Rasch simple logistic model. Minneapolis: University of Minnesota, TR No. 3006, 1972. Investigated the use of G. Rasch's simple logistic model in obtaining test-free ability estimates. Tests employing word, picture, symbol, and number analogies were administered to college and high school students. Results show that the Rasch model was not an improvement over percentile ranks in estimating individual ability. These results, however, were an artifact of the research design. Shortcomings in the present design, as well as in the one used by B. Wright are discussed, and a more appropriate design is suggested. (Roberts)

Tinsley, H. E., & Dawis, R. V. An investigation of the Rasch simple logistic model: Sample-free item and test calibration. Minneapolis, MN: U. Minnesota, TR No. 3005, 1972, Intestigated the use of G. Rasch's 1960 simple logistic model in item and test calibration. Tests employing word, picture, symbol, and number analogies were administered to 1,404 college students, 484 high school students, 289 civil service clerical employees, and 90 vocational rehabilitation clients. Results suggest that Rasch item-easiness estimates are invariant with respect to the ability of the calibrating sample when an adequate sample is employed. The invariance of the Rasch item-easiness estimates was related to the goodness-of-fit of the items to the Rasch model. The deletion of items with low Rasch probabilities increased the invariance of the Rasch item-easiness estimates. Estimates of the amount of ability indicated by the raw scores on a test (ability estimates) were also invariant with respect to the ability of the calibrating sample for tests of 25 or more items, evan when relatively small samples were employed. (Roberts)

Tinsley, H. E., & Dawis, R. V. An investigation of the Rasch simple logistic model: Sample free item and test calibration. Educational and Psychological Measurement, 1975, 35, 325-339. Administered tests employing word, picture, symbol, and number analogies to high school students, undergraduates, clerical employees, and clients of the Minnesota Division of Vocational Rehabilitation. Use of the simple logistic model for item and test calibration indicated that Rasch item easiness ratios and z item difficulty ratios were invariant

with respect to the ability of the calibrating sample when an adequate sample was employed and the test design did not incorporate biasing factors. The invariance of the Rasch item easiness ratios was related to the goodness-of-fit of the items to the Rasch model in that the deletion of items with low Rasch probabilities increased the invariance of the Rasch item easiness ratios. Estimation of the amount of ability indicated by the raw scores on a test was invariant with respect to the ability of the calibrating sample for tests of 25 or more items, even when samples of fewer than 100 Ss were studied. (Roberts)

Toman, W. Experiments with field data. Psychologie in Erziehung und Unterricht, 1976, 23(2), 80-90.

Uhl, N. P., & Briley, T. S. Application of the Rasch Model to a measure of environmental perception. Paper presented at the meeting of the Northeastern Educational Research Association, Grossingers, New York, November 1970. (Forster, Passmore)

Urry, V. W. A Monte Carlo investigation of logistic test models. (Doctoral dissertation, Purdue University, 1970.) Dissertation Abstracts International, 1971, 31, 6319B. (University Microfilms No. 71-9475) (Forster, Passmore)

Urry, V. W. Approximations to item parameters of mental test models and their uses. Educational and Psychological Measurement, 1974, 34, 253-269.

Urry, V. W. A five-year quest: Is computerized adaptive testing feasible? Proceedings of the First Conference on Computerized Adaptive Testing (PS-75-6, Personnel Research and Development Center, U. S. Civil Service Commission). Washington, DC, U. S. Government Printing Office, 1976. (Superintendent of Documents Stock No. 006-000-00940-9)

Urry, V. W. Tailored testing: a successful application of latent trait theory. Journal of Educational Measurement, 1977, 14(2), 181-196. This paper describes tailored testing, a successful application of latent trait theory. It is shown that tailored testing can be economically applied to practical problems in educational measurement. Necessary conditions for the development of an efficient system of tailored testing are discussed. Applications of tailored testing for single and multiple abilities are illustrated. The U. S. Civil Service Commission's plans for tailored testing in an employment setting are presented. The major psychometric and economic problems in tailored testing have been overcome. Tailored testing requires the use of far fewer items than does comparable paper-and-pencil testing. Thus, the technological costs of tailored testing are no longer prohibitive. To date, tailored testing has been used only with ability tests. Further applications in areas such as guidance and counseling, knowledge and achievement testing, and computer-managed instruction are described. The implications of tailored testing for the training of future measurement theorists and practitioners are discussed.

Urry, V. W. Tailored testing: A spectacular success for latent trait theory. Springfield, VA: National Technical Information Service, 1977.

Vogt, D. K. An extension of the Rasch model to the case of polychotomously scored items. (Doctoral dissertation, University of Maryland, 1971). Dissertation Abstracts International, 32(3-A), 1203. (Forster, Passmore)

Vogt, D. K. On an extension of the Rasch model to the case of polychotomously scored items. Paper presented at the meeting of the American Educational Research Association, New York, NY, February 1971, ED 049 289. (Forster, Passmore)

Valgeirsdottir, S. T. Adaptive test construction using the Rasch model of measurement. (Doctoral dissertation State University of New York, Buffalo, 1075) Dissertation Abstracts International, 1975, 35(10-A) 6522.

Volavka, J., Mednick, S. A., Sergeant, J., & Rasmussen, L. Electroencephalograms of XYY and XXY men. British Journal of Psychiatry, 1977, 130(1), 43-47.

Waller, M. I. Removing the effects of random guessing from latent trait ability estimates. Research Bulletin 74-32. Princeton, NJ: Educational Testing Service, 1974. In latent trait models the standard procedure for handling the problem caused by guessing on multiple choice tests is to estimate a parameter which is intended to measure the "guessingness" in-herent in an item. Birnbaum's three parameter model, which handles guessing in this manner, ignores individual differences in guessing tendency. This paper presents a mode- or procedure which uses the information contained in the interaction between a person and an item to remove the effects of random guessing from estimates of ability, difficulty, and discrimination. Simulated and real data are presented which support the model in terms of fit and information.

Waller, M. I. Estimating guessing tendency. Research Bulletin 74-33. Princeton, NJ: Educational Testing Service, 1974.

Waller, M. I. Estimating parameters in the Rasch model: Removing the effects of random guessing. Research Bulletin 76-8. Princeton, NJ: Educational Testing Service, 1976. A method of estimating the parameters of the Rasch model removing the effect of random guessing is presented. The procedure is an application of the ARRG (Abilities Removing Random Guessing) model recently developed for two parameter latent trait models. Under the Rasch model ARRG provides for estimation of abilities, removing the effects of random guessing, without requiring the use of a computer. Monte Carlo simulations are employed to examine the accuracy of the resulting ability estimates.

Weber, M. B. An examination of the bilevel dimensionality of lower and higher mental processes in probability achievement. Educational and Psychological Measurement, 1977, 37(4), 987-990. Bilevel dimensionality of probability was examined via factor analysis, Rasch latent trait analysis, and classical item analysis. Results suggest that when non-standardized measures are the criteria for achievement, relying solely on estimates of content validity may lead to erroneous interpretation of test score data.

Weiss, D. J. Strategies of adaptive ability measurement. Research Report 74-5. Minneapolis, MN: Psychometric Methods Program, Department of Psychology, University of Minnesota, 1974.

Weiss, D. J. Adaptive testing research at Minnesota: Overview, recent results, and future directions. In C. L. Clark (Ed.), Proceedings of the First Conference on Computerized Adaptive Testing. Washington, DC: United States Civil Service Commission, 1976.

Weiss, D. J. Computerized ability testing, 1972-1975. Final Report. Minneapolis, MN: University of Minnesota, Department of Psychology, Psychometric Methods Program, April 1976.

Weiss, D. J. Applications of computerized adaptive testing, D. J. Weiss, Ed. Research Report 77-1. Proceedings of a Symposium presented at the Annual Convention of the Military Testing Association, October 1976. Minneapolis, MN: Minnesota University, Department of Psychology. This symposium consists of five papers and presents some recent developments in adaptive testing which have applications to several military testing problems. The overview, by J. R. McBride, defines adaptive testing and discusses some of its item selection and scoring strategies. Item response theory, or item characteristic curve theory, is also described. In the second paper, J. B. Sympson explicates the role of latent trait theory in measurement for criterion prediction and in criterion referenced measurement. C. D. Vale then discusses the use of adaptive testing procedures to make ability classification decisions (i.e., cutting score decisions). In the fourth paper S. M. Pine argues that a major problem in current efforts to develop less biased tests is an over-reliance on classical test theory. Item characteristic curve theory is offered as a more appropriate measurement model. In the final paper by I. I. Bejar, two relatively recent developments in psychometric theory, the assessment of partial knowledge and research in adaptive testing, are reviewed.

Wells, R. A. The probabilistic interpretation of test scores calibrated by the Rasch Model (Doctoral dissertation, University of Georgia, 1973). Dissertation Abstracts International, 1974, 34(07-A), 4012. (University Microfilms No. 73-31977)

White, P. O. A note on Keats' generalization of the Rasch model. Psychometrika, 1976, 41, 405-407. In J. A. Keats's paper, he discusses the general form of the projective transformation and applies it to G. Rasch's (1960) problem-solving model. In the present note, an alternative derivation is given for a simple test model which incorporates an ability parameter for the S and difficulty level and guessing parameters for the problem. The probability of a correct response to the problem is a projective transformation of the problem difficulty. The ability and difficulty parameters separate into additive components. (Roberts)

Whitely, S. E. Solving verbal analogies - some cognitive components of intelligence-test items. Journal of Educational Psychology, 1976, 68(2), 234-242.

Whitely, S. E. Models, meanings and misunderstandings: Some issues in applying Rasch's theory. Journal of Educational Measurement, 1977, 14(3), 227-235. A debate concerning specific issues and the general usefulness of the Rasch latent trait model is continued. Methods of estimation, necessary sample size, and the applicability of the model are discussed.

Whitely, S. E., & Dawis, R. V. The nature of objectivity with the Rasch Model. Journal of Educational Measurement, 1974, 11, 163-178. Discusses the claim that the G. Rasch model leads to a higher degree of objectivity in measurement than has been previously possible. Population-invariant item and ability calibrations along with the statistical equivalency of any 2 item subsets are supposedly possible if the item pool has been calibrated by the Rasch model. An analysis is made of the conditions under which claims of objectivity will be substantiated, with special emphasis on the nature of equivalent forms. It is concluded that the real advantages of the Rasch model will not be apparent until the technology of trait measurement becomes more sophisticated. (Roberts)

Whiteley, S. E., & Dawis, R. V. The influence of test context on item difficulty. Educational & Psychological Measurement, 1976, 36(2), 329-337. Item analysis techniques commonly assume that item difficulties are stable over varying test contexts. The current emphasis on information-efficiency and individualized selection of items has made context stability a critical assumption. The simple item difficulties and Rasch log easiness values of 15 verbal analogy items were compared over 7 internally consistent analogy tests. Results indicate that 60% of the items varied over contexts and that content difficulty was a factor for 27% of the items. The general applicability of item-parameter-invariant models is questioned by the finding of contextual interaction for parameter estimates. (Roberts)

Wiebel, B., & Torne, J. V. Neuroticism and extraversion in auto-aggressive children. Monatsschrift fur Kinderheilkunde, 1977, 125(5), 589.

Wiedel, K. H., & Carlson, J. S. Factorial structure of Raven Colored Progressive Matrices Test. Educational and Psychological Measurement, 1976, 36(2), 409-413.

Williams, E. A. The stability of item characteristic curve parameters across groups. (Doctoral dissertation, University of Illinois at Champaign-Urbana). Ann Arbor, MI: University Microfilms, 1971, No. 72-12, 437. (Forster, Passmore)

Willmott, A. Asessment and performance. Oxford Review of Education, 1978, 4(1), 51-64.

Willmott, A. S., & Fowles, D. E. The objective interpretation of test performance: The Rasch Model applied. Atlantic Highlands, NJ: NFER Publishing Co., 1974.

Wood, R. L. Response-contingent testing. Review of Educational Research, 1973, 43(4), 529-544. Presents a longitudinal review which considers response-contingent testing in its developmental aspects from early work in this area to the contemporary activity dating from the last decade and into the present. Response-contingent testing is discussed in terms of its contribution to decision-making. Tailored testing and latent trait estimation as well as their limitatins are considered.

Wood, R. L. Fitting the Rasch model - a heady tale. British Journal of Mathematical and Statistical Psychology, 1978, 31, 27-31.(Roberts)

Wood, R. Trait measurement and item banks. In D. DeGruijter and L. T. van der Kamp (Eds.), Advances in Psychological and Educational Measurement. London: Wiley, 1976.

Wood, R. L, & Lord, F. M. A user's guide to LOGIST. Research Memorandum 76-4. Princeton, NJ: Educational Testing Service, 1976.

Wood, R. L, Wingersky, M. S., & Lord, F. M. LOGIST: A computer program for estimating examinee ability and item characteristic curve parameters. Research Memorandum 76-6. Princeton, NJ: Educational Testing Service, 1976.

Woodbury, M. A., & Novick, M. R. Maximizing the validity of a test battery as a function of relative test lengths for a fixed total testing time. Journal of Mathematical Psychology, 1968, 5, 242-259.

Woodcock, R. W. Woodcock Reading Mastery Tests, Circle Pines, MN; American Guidance Service, 1974.

Woodcock, R. W., & Dahl, M. N. A common scale for the measurement of person ability and test item difficulty, (AGS paper #10). Circle Pines, MN: American Guidance Service, April 1971. (Forster, Passmore. Roberts)

Wright, B. D. Sample-free test calibration and person measurement. In Proceedings of the 1967 Invitational Conference on Testing Problems. Princeton, NJ: Educational Testing Service, 1968. (Forster, Passmore)

Wright, B. D. Mean value estimation. A paper presented at an American Educational Research Association pre-session on Person-free Item Calibration and Item-free Person Measurement, Minneapolis, MN, 1970. (Forster, Passmore)

Wright, B. D. Crude estimation. Paper presented at an American Educational Research Association pre-session on Person-free Item Calibration and Itrem-free Person Measurement, Minneapolis, MN, 1970. (Forster, Passmore)

Wright, B. D. A simple method for organizing and evaluating the objective measurement of a latent trait presented in the language of a short, convenient FORTRAN program. Mini Research Training Course, Annual Meeting of the American Educational Educational Research Association, New Orleans, 1973. (Forster, Passmore)

Wright, B. D. Solving measurement problems with the Rasch Model. Journal of Educational Measurement, 1977, 14, 97-116. Explains the Rasch model for sample-free item analysis and test-free person measurement. Methods are provided for estimating model parameters from data and for evaluating the statistical fit of these estimates to the data. The Rasch model is compared with other latent trait models, and it is shown why attempts to estimate item discrimination must run into trouble. The paper describes (a) how to build calibrated item banks within which all possible tests are automatically equated to a common scale, (b) how to criterion reference and norm this scale, (c) how to design "bad tests" for optimal person measurment, and (d) how to evaluate the statistical fit and precision of each person's measure.

Wright, B. D. Misunderstanding the Rasch Model. Journal of Educational Measurement, 1977, 14, 219-225. Statements made in a previous article of this journal concerning the Rasch latent trait test model are questioned. Methods of estimation, necessary sample sizes, several formuli, and the general usefulness of the Rasch model are discussed.

*Wright, B. D. The Rasch Model for test construction and person measurement. Paper presented at the Fifth Annual Conference and Exhibition on Measurement and Evaluation, Office of the Los Angeles County Superintendent of Schools, Division of Program Evaluation, Research and Pupil Services, Los Angeles, CA, March 1978.

Wright, B. D., & Douglas, G. A. Best test design and self-tailored testing Research Memorandum No. 19. Chicago, IL: University of Chicago, Department of Education, Statistical Laboratory, 1975.

Wright, B. D., & Douglas, G. A. Better procedures for sample-free item analysis. Research Memorandum No. 20. Chicago IL: University of Chicago, Department of Education, Statistical Laboratory, 1976.

Wright, B. D., & Douglas, G. A. Rasch item analysis by hand. Research Memorandum No. 21. Chicago, IL: University of Chicago, Department of Education, Statistical Laboratory, 1976.

Wright, B. D., & Douglas, G. A. Best procedures for sample-free item analysis. Applied Psychological Measurement, 1977, 1(2), 281-295.

Wright, B. D., & Douglas, G. A. Conditional versus unconditional procedures for sample-free item analysis. Educational and Psychological Measurement, 1977, 37, 573-586. Two procedures for Rasch, sample-free calibration are reviewed and compared for accuracy. The theoretically ideal "conditional" procedure is impractical for more than fifteen items. The more practical but biased "unconditional" procedure is discussed in detail.

Wright, B. D., & Mead, R. J. BICAL: Calibrating rating scales with the Rasch Model. Research Memorandum No. 23A. Chicago, IL: University of Chicago, Department of Education, Statistical Laboratory, 1976.

Wright, B. D., & Mead, R. J. CALFIT: Sample-free item calibration with a Rasch measurement model. Research Memorandum No. 18. Chicago, IL: University of Chicago, Department of Education, Statistical Laboratory, 1976.

Wright, B. D., & Mead, R. J. The use of measurement models in the definition and application of social science variables. (Grant DAHC19-76-G-0011, Final Report). U. S. Army Research Institute for the Behavioral and Social Sciences, Arlington, VA, June, 1977.

Wright, B. D., Mead, R. J., & Draba, R. Detecting and correcting test item bias with a logistic response mode. Research Memorandum No. 22. Chicago, IL: University of Chicago, Department of Education, Statistical Laboratory, 1976.

Wright, B. D., & Panchapakesan, N. A procedure for sample-free item analysis. Educational and Psychological Measurement, 1969, 29, 23-48. (Forster, Passmore)

Wright, B. D., & Panchapakesan, N. Sample free measurement: Part 3, The probability of binomial observations. Paper presented at an American Educational Research Association pre-session on Person-free Item Calibration and Item-free Person Measurement, Los Angeles, 1969. (Forster, Passmore)

Wright, B. D., & Panchapakesan, N. Sample free measurement: Part 4, The estimation of parameters. Paper presented at an American Educational Research Association pre-session on Person-free Item Calibration and Item-free Person Measurement, Los Angeles, 1969. (Forster, Passmore)

Wright, B. D., & Panchapakesan, N. Item analysis by the Rasch model, UCSL801. University of Chicago Computation Center, Social Science Program Library, 1970. (Forster, Passmore)

Wright, B. D., & Skirmont, R. MESAMAX, Computer Program, 1972. (Available from Ronald Skirmont, CSIP, Inc., 1525 E. 53rd St, Chicago, IL 60615).

Wright, B. D., & Stone, M. H. Best test design: A handbook for Rasch measurement. Palo Alto: Scientific Press, 1978.